A STUDY IN *Sorghum bicolor*: QTL ANALYSIS OF PHOTOPERIOD SENSITIVE SORGHUMS, EVALUATION OF SORGHUM × SUGARCANE HYBRIDS AND TRAIT INTROGRESSION FOR INTERGENERIC HYBRID IMPROVEMENT

A Thesis

by

MATTHEW SCOTT BARTEK

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee, William Rooney
Committee Members, Patricia Klein
David Stelly
John Mullet
Jorge da Silva
Head of Department, David Baltensperger

August 2013

Major Subject: Plant Breeding

ABSTRACT

Recently designated as a bioenergy crop, Sorghum is rather unique as it can produce large quantities of cellulose or sugar which can be used to produce advanced biofuels or compounds. Sweet sorghum contains high levels of sugars and biomass sorghums consist primarily of ligno-cellulosic biomass. Improvement of both sorghum types is essential for maximizing production and conversion efficiency. Photoperiod sensitive sorghum is thought to maximize biomass production yet maturity influence on biomass production and composition is not fully understood. Utilizing sorghum for sugar production has increased efforts to develop sweet sorghums with sugar yields similar to sugarcane. Hybridization of these species has been investigated with, until recently, little success. Testing newly developed intergeneric hybrids and improvement of parents used in their creation will determine their feasibility and improve hybrid performance.

Objectives of this research are multifaceted. First, analyze photoperiod sensitive sorghum in varying day length environments to determine maturity effects on plant phenotype, composition, and QTL detection. Second, analyze intergeneric sorghum × sugarcane hybrids to determine agronomic performance in relation to sugarcane. Lastly, introgress the *iap* allele into sweet sorghum females for use in intergeneric hybrid creation.

Photoperiod sensitive sorghum RILs were evaluated in College Station and Weslaco, Texas and Puerto Rico which caused differential expression of plant maturity. Genetic control of trait expression was high for each location. Results indicate gradual induction of

plant maturity increases detection of phenotypic QTL and detection of compositional QTL increases when maturity effects on plant phenotype are reduced.

Intergeneric sorghum × sugarcane F_1 hybrids were compared to sugarcane in Weslaco, Texas in 2011. Each hybrid expressed agronomic traits similar or better than that of the sugarcane variety. High levels of repeatability and genetic influence on trait expression were observed. Overall performance of the sugarcane variety was better than any individual hybrid tested.

Introgression of *iap* into sweet sorghum was successful and generated seventeen new sweet sorghum female genotypes possessing the allele. Only two genotypes exhibited higher brix readings and both were later maturing than Tx3361. Height and maturity of all developed genotypes varied and desirability of developed lines was similar to Tx3361.

ACKNOWLEDGMENTS

In dedication to my beautiful wife Cara and daughter Caroline, without you none of this would be possible. Your support and encouragement helped me push through the difficult process of graduate school. The long hours and research trips have been especially tough on the two of you. I love you both and you make me proud to be a husband and father.

Thanks to Dr. Bill Rooney for taking a chance on a turf student with less than impressive GRE scores. Your mentorship and guidance through my development as a scientist has helped ease the transition from student to professional. I would also like to thank my entire graduate committee; Dr. Trish Klein, Dr. David Stelly and Dr. John Mullet for your continued help with genetic and cytogenetic training throughout my professional development. Your knowledge has helped me become a better breeder and more complete scientist.

Finally I would like to thank the graduate and undergraduate students in the Texas A&M Sorghum Breeding lab; Dr. Leo Hoffman Jr., Dr. Jason Anderson, John Gill, Dustin Herb, Payne Burkes, Terry Felderhoff, Steve Labar, Miguel Gutierrez, Dustin Borden, Michael Klepac, Josh Herrington, and Catherine Lettunich. I apologize if I left anyone out. You guys truly keep things moving forward and no project would be completed without you. My project required many harvest and extensive amounts of field and greenhouse work. Your long hours and commitment to the sorghum program was greatly appreciated.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER I

INTRODUCTION

Increasing worldwide demand for food and fuel has placed new emphasis on creating cultivated crop species with higher yields and improved tolerances to biotic and abiotic stresses (Maqbool et al., 2001; Mathews et al., 2000; Dillon et al., 2007). Three species that have been identified as bioenergy crops are sorghum, sugarcane, and corn (Ahn and Tanksley 1993; Giussani et al., 2001; Paterson 2008; Paterson et al., 2009). Conversion of corn grain into ethanol however is not viewed as a sustainable biofuel production model due to its limited output of ethanol yield and direct competition with food production (Rooney et al., 2007). This limitation of corn-based ethanol allows a window of opportunity for the study and production of sorghum and sugarcane as sustainable biofuel feedstocks within the U.S.

Sorghum (*S. bicolor*) is the world's fifth most important grain crop based on production and is second as a source of U.S. grain-based ethanol (Paterson, 2008). Drought and heat stress tolerance in sorghum are valued in current production regions and ensure its continued production. The importance of these traits is becoming more evident as increases in demand for food force agriculture to be more efficient with regards to water, fertilizer and land use (Paterson, 2008).

Sorghum domestication began in Ethiopia around 4000-3000 B.C. and spread to surrounding areas, resulting in new, locally adapted varieties (Dillon et al., 2007). Human mobility and cultural trade routes resulted in improved lines spreading throughout the world and eventually America in the late 1800s to early 1900s (Dillon et al., 2007). As sorghum migrated across the globe, local selection resulted in significant genetic diversity within the

species (Wright, 1990). Analysis of sorghum genetic diversity indicates considerable polymorphism across the species which has been underexploited for crop improvement (Wu et al., 2004; Abu Assar et al., 2005; Deu et al., 2006; Kayode et al., 2006; Dillon et al., 2007).

Germplasm diversity of crop species has been the basis for crop improvement and an integral tool to plant breeders. Today, most of the sorghums in the world collection are photoperiod sensitive, meaning that reproductive growth is initiated once day lengths are sufficiently reduced to meet the required short day photoperiod (Reddy et al., 2006). Beginning in the late 1960s, photoperiod sensitive sorghum genotypes were being converted to photoperiod insensitive for their use in U.S. breeding programs (Stephens et al., 1967). The sorghum conversion program has provided sorghum improvement programs in temperate environments access to genotypes that have been used to improve biotic and abiotic stress resistance as well as improved yield and quality for both grain and forage (Stephens et al., 1967). This progress was achieved even though the majority of the sorghum collections are still available only as photoperiod sensitive versions (Rosenow and Dahlberg 2000).

Recently, sorghum was designated as a bioenergy crop by the U.S. Department of Energy (Perlack et al., 2005. Among the bioenergy crops, sorghum is rather unique in that it can be used to produce large quantities of carbohydrates as either starch (grain sorghum), sugar (sweet sorghum) or cellulose (biomass sorghum) which can then be used to produce advanced biofuels or compounds. Each of these different sorghum types maximizes one category of carbohydrates but all generally produce measurable quantities of all three. Sweet sorghum contains high levels of sugars and biomass sorghums consist primarily of ligno-cellulosic biomass but both are dedicated bioenergy feedstocks (Rooney et al., 2007). While

sweet sorghum can be processed in the same way as sugarcane to produce ethanol from sugar, the use of structural carbohydrates relies on conversion technologies that are still being developed (Rooney et al., 2007). Effectiveness and efficiency of biomass conversion has been shown to be affected by biomass composition (Monti et al., 2008).

To maximize the rate of genetic improvement, genomic research to detect QTL and the genes underlying both biomass yield and composition has begun (Felderhoff et al., 2012, Murray et al., 2008; Ritter et al., 2008; Lin et al., 1995). In addition to finding unique QTL for biomass yield and composition, these studies have also concluded that genes controlling maturity and height have as much influence, if not more, on these traits. This implies that maturity and height, which are both highly heritable, can be rapidly selected to maximize productivity and composition to differing maturity adaptation zones. The results also imply that to fully identify additional loci underlying biomass yield and composition, eliminating the effects due to these loci will be critical (Corn 2009).

Approaches to modify maturity in biomass sorghum have been developed. Genes controlling both maturity *per se* and the response to photoperiod have been identified, characterized and in some cases, cloned (Childs et al., 1997; Murphy et al., 2011). The use of photoperiod sensitive sorghums with prolonged periods of vegetative growth significantly increases biomass yields (Rooney et al., 2007). While the effect of photoperiod sensitivity is obvious from previous research, there have been no studies to evaluate the genetic variation for biomass yield and composition in segregating families that are wholly photoperiod sensitive. Given the phenotypic variation in long-day environments, the interaction of these populations in environments of varying and short day lengths is essential (Rooney et al., 2007).

The relatively small genome size of sorghum (~730Mb) (Paterson et al., 2009), combined with a high level of genetic diversity within the species, makes it an attractive candidate for genome exploration and exploitation toward crop improvement for bioenergy (Dillon et al., 2007; Paterson et al., 2009). Sorghum genome mapping has resulted in high density genetic maps useful for the study and analysis of genes controlling traits of agronomic importance (Bowers et al., 2003; Menz et al., 2002; Feltus et al., 2004). Information gleaned from these studies can assist in introgression of specific traits into elite germplasm (Menz et al., 2002. Even though such studies have identified genetic locations of many important quantitative traits (QTL) and increased our understanding of them, exploitation of these traits is still limited (Bernardo 2008).

Quantitative trait loci mapping increases our biological understanding of species while identifying markers, which can be useful to breeders during trait selection (Bernardo 2008). From a plant breeding perspective, QTL mapping is divided into separate yet equally important goals. First, QTL mapping can identify major candidate traits for introgression, and second, identification of traits can lead to development of genetic markers useful in marker assisted selection (MAS) (Bernardo 2008). Finally, this information can be used to clone the gene if this is a desired or necessary goal. Several studies have identified QTL that influence both biomass yield and composition (Felderhoff et al., 2012, Murray et al., 2008; Ritter et al., 2008; Lin et al., 1995) and some of those identified have been found in multiple studies indicating increased reliability of identification. For example, Murray et al. (2008) and Felderhoff et al. (2012) both identified a QTL for brix on sorghum chromosome 3. Two QTL for height were also found to be on chromosome 6 in separate studies conducted by Ritter et al. (2008) and Felderhoff et al. (2012). Should phenotyping methods of populations

used in QTL identification become more uniform, identification of co-linear QTL will increase overall reliability and allow for identification of truly novel QTL from different populations. Marker development and incorporation into breeding programs of these identified traits has the potential to improve the yield and quality of sorghum biomass

In recent decades introduction of traits which improve plant fitness have allowed sorghum to become adapted to more diverse growing conditions than previously available while simultaneously increasing grain yield as well as biotic and abiotic stress tolerance (Maqbool et al., 2001).

Improvement of all crop species including sorghum has long relied on hybridization within the species (Kuhlman et al., 2008; Hodnett et al., 2010). While less frequent, interspecific and even intergeneric hybridization has produced breakthrough improvements and innovations in crops. Haploid induction of wheat following pollination by maize (Laurie and Bennet, 1988) produced great promise towards introgression of maize DNA into wheat as well as a valuable alternative to wheat (*Triticum* spp. L.) × barley (*Hordeum bulbosum* L.) haploid induction. Similarly, the production of Triticale (*Triticale hexaploide* Lart.) through hybridization of wheat and rye (*Secale cereal* L.) allowed the expression of desirable genes from both species in the hybrid plant (Zenkteler and Nitzsche, 1984). In addition to these innovative advancements, wide hybridization is a means to enhance genetic gains through the introgression of genetic diversity that is not readily available in the primary gene pool of cultivated crops (Sharma, 1995) or improve upon currently utilized breeding methodologies. Intergeneric hybridization within Poaceae may provide opportunities for trait introgression between genera and assist in improving high value crops within the grass family (Hodnett et al., 2010).

Sugarcane (*Saccharum* spp. L.), like sorghum, has had only a portion of the available genetic diversity available incorporated or introgressed into domesticated varieties despite its high value as an agricultural product (Dillon et al., 2007). Sugarcane ranks first in global production as a sugar producing crop, accounting for 60% of raw sugar consumed worldwide, and is grown in more than ninety countries worldwide (Grivet and Arruda 2001; Dillon et al., 2007). A significant amount of sugarcane is now used for ethanol production, especially in Brazil (Goldemberg et al., 2004). In relation to the economic importance of sugarcane, genomic research within the species has lagged behind other agronomically important crops, presumably due to the complexity of its genome (Grivet and Arruda, 2001).

The evolutionary divergence of sorghum and sugarcane is estimated to have occurred roughly 5 million years ago making the two close relatives among the cultivated crop species (Paterson et al., 2004; Dillon et al., 2007). Guimareaes et al. (1997) found nearly perfect marker co-linearity between Paupan *Saccharum* and *Sorghum* through comparative mapping of the two genomes. Four cases of marker order change due to inversion were found in comparisons between *Sorghum* and *S. robustum* while no marker order changes were seen in comparison to *S. officinarum* (Guimaraes, 1997). The co-linearity between *Sorghum* and *Saccharum* indicate a strong conservation between the two genera (Guimaraes 1997).

Attempts to introgress traits between *Saccharum* and sorghum began in 1929 when Venkatram attempted to produce early maturing sugarcane (Thomas and Venkatraman 1930; Gupta et al., 1978). Further efforts were made toward the production of hybrids in the 1970s when tropical sorghum breeders sought to introgress shoot-fly resistance from sugarcane into sorghum (Young 1972; Gupta et al., 1978). Gupta et al. (1978) reported successful production of *Saccharum* × *Sorghum* hybrids but ultimately trait introgression could not be

established. Nair (1999), however, confirmed the first successful *Sorghum* × *Saccharum* hybrid when five hybrid seedlings were recovered from 3,670 pollinations. Unfortunately, the hybrids produced were of limited breeding value.

Limited production of *Sorghum* × *Saccharum* hybrids are the result of reproductive barriers and removing or overcoming these barriers is necessary in achieving intergeneric hybridization (Hodnett 2005). The discovery of the *S. bicolor* mutant allele *iap* (Inhibition of Alien Pollen) has been shown to remove or reduce some of these reproductive barriers (Laurie and Bennett 1989; Hodnett et al., 2005; Price et al., 2006; Kuhlman 2008; Bartek et al., 2012) and has led to the development of sorghum genotype Tx3661 (Kuhlman and Rooney 2011). Sorghum genotype Tx3361 has facilitated the production of *Sorghum* × *Saccharum* intergeneric hybrids at levels not previously seen (Hodnett et al., 2010). Hodnett et al. (2010) produced 14,141 seed from 252 Tx3361 × *Saccharum* crosses over a three year period. Seed set was as high as 53% for a single *Saccharum* pollinator and successful hybridization appeared to be male genotype specific (Hodnett et al., 2010). While these F_1 hybrids did not produce viable seed, the ability to produce large quantities of intergeneric *Sorghum* × *Saccharum* hybrids provides new opportunities for genetic improvement in both species (Hodnett et al., 2010). The utility of Tx3361 in hybrid combination with sugarcane may be maximized through development of improved Tx3361 parental genotypes.

With existing high levels of genetic diversity between the species and potentially large numbers of intergeneric hybrids, segregation within new *Sorghum* × *Saccharum* populations should allow selection of elite and novel germplasm (Hodnett et al., 2010). Screening of F_1 hybrids for desirable agronomic traits may provide opportunities for novel hybrid crop development (Hodnett et al., 2010). Production of *Sorghum* × *Saccharum*

hybrids with the ability to accumulate sugar from sugarcane with enhanced water-use efficiency from sorghum has potential to provide significant value to producers and breeders (Hodnett et al., 2010).

It is widely accepted that sorghum contains direct potential as a bioenergy feedstock (Stefaniak et al., 2012). Continued genetic research towards improvement of the species is fundamental for production of highly adapted genotypes with biomass composition desirable for use in emerging energy markets (Rooney et al., 2007), Dedication to this effort will allow production of superior genotypes containing improved agronomic and compositional traits (Stefaniak et al., 2012). With the reality that bioenergy feedstocks will be comprised of many different crop species across varying environments, it is necessary to examine the benefits and potential of sorghum to fill that need (Rooney et al., 2007).

The objectives of this research were:

1. Analyze a photoperiod sensitive sorghum RIL population to identify QTL for biomass yield and composition in long and short day environments.

2. Evaluate agronomic performance and composition of *Sorghum* × *Saccharum* F_1 hybrids.

3. Develop sorghum breeding lines homozygous for the *iap* allele that possess drought tolerance and/or high stalk sugar content.

CHAPTER II

IDENTIFICATION OF PHENOTYPIC QTL IN A PHOTOPERIOD SENSITIVE RIL SORGHUM POPULATION

Introduction

Sorghum [*Sorghum bicolor* (L. Moench)] is an important grain and forage crop throughout the world. Because of its importance, there have been systematic breeding efforts to improve the crop for over a century (Rooney 2000). Genotypes are improved through the accumulation of desirable alleles through selection. Regardless of whether genotypes are bred for use as pure line cultivars or as hybrids, the focus of breeding programs is on yield, adaptation and quality (Rooney 2004)

In the past five years, there has been significant interest and funding to develop and deploy energy crops. For several reasons, sorghum has been identified as one species that can be used in this capacity. First, sorghum is a highly productive grass that uses NADP-ME C4 photosynthesis for carbon assimilation. Sorghum is a low risk annual hybrid amenable to normal crop rotations that maintain soil fertility, reduce pest pressures, and replace perennial crops when stands are unexpectedly lost. Under optimum irrigated growth conditions, current energy sorghum hybrids have the potential to produce large quantities of biomass per acre (Rooney et al., 2007). Sorghum has excellent drought tolerance and high water use efficiency, critical attributes for bioenergy crop production in environments where irrigation is not available, too expensive, or depletes water reserves. Sorghum is widely adapted and highly amenable to U.S. production and cultivation systems. In more southern regions of the U.S., the regrowth potential of sorghum may be important for ratoon crop production and to reduce soil erosion. Unlike several other proposed energy crops, extensive cultivation worldwide is supported by numerous breeding programs and an extensive

germplasm collection of ~40,000 accessions that includes useful variation for an array of bioenergy traits including biomass yield, composition, and drought tolerance.

In sorghum, genetic mapping and QTL analysis have resulted in high density genetic maps useful for the identification and study of quantitative trait loci (QTL) (Menz et al., 2002. In addition, the completed genome sequence of sorghum makes it possible to fine map and potentially clone genes of importance for most agronomic traits (Paterson et al., 2009). Multiple studies in sorghum have identified QTL that influence biomass yield and composition (Felderhoff et al., 2012, Murray et al., 2008; Ritter et al., 2008). Most of these studies have concluded that these traits are strongly influenced by both maturity and height. Felderhoff et al. (2012) found positive correlations of both height and flowering time on brix and harvest yield. Lin et al. (1995) reported that two of the three QTL for maturity were directly associated with QTL for height. Significant correlations between biomass yield and flowering time were also seen by Ritter et al. (2008) as well as between height and days to flowering. In that study, six QTL for height were found across four linkage groups and five QTL for flowering time were spread across four linkage groups. Murray et al.,(2008) found two QTL for height and two QTL for flowering time across three linkage groups with at least one QTL for each trait co-localizing to a single linkage group. The mapping populations utilized in these studies varied greatly in the generation of individuals being mapped (F_2 - F_6), number of individuals within the population (176- 370), and parental phenotypes (tall, late flowering × short, early flowering). This resulted in populations with high levels of phenotypic variances and elevated heritability estimates. As expected, different QTL were detected across these populations due to QTL segregating in some populations but not others as well as variation in the environments in which they were evaluated.

While the previously mentioned studies have focused on the use of bi-parental mapping populations for QTL identification, genome wide association mapping (GWAS) has recently been implemented in sorghum to identify the genetic basis of agronomic traits (Morris et al., 2013). Morris et al. (2013) used a diverse association panel of 971 accessions collected from world germplasm collections. By scanning the genomes of each individual within the panel using single nucleotide polymorphic (SNP) markers, they were able to identify patterns of sorghum diffusion from its center of origin and identify sources of genes commonly used in breeding programs (Morris et al., 2013). Results of their study indicate that GWAS is a new and effective means for identifying QTL for traits expressing low levels of genetic diversity. This is a promising approach for QTL studies in photoperiod sorghums which maintain high levels of diversity for height and maturity on chromosomes 6, 7 and 9. Chromosome 6 is most affected by conversion to photoperiod insensitivity where most of the chromosome has been introgressed from the same donor (BTx406) during the conversion process (Morris et al., 2013).

To date, QTL mapping efforts for bioenergy traits have utilized sorghum populations that segregate for both height and maturity. Given that these traits are pleiotropic for biomass productivity, they exert a large effect and may mask additional QTL for quality traits that are unrelated to maturity and/or height (Murray et al., 2008). Thus, there is a need to identify QTL influencing biomass yield independent of the effects of maturity and plant height.

Recent approaches to increase biomass production of energy sorghums are to prolong vegetative growth through delaying flowering time (Murphy et al., 2011). Wide ranges of photoperiod sensitivity within sorghum have produced much variation for maturity responses

and flowering times in the crop (Pittendrigh and Minis 1964; Quinby 1974). Historically, maturity of sorghum was found to be controlled by four maturity loci designated *Ma1*, *Ma2*, *Ma3*, and *Ma4*, with *Ma1* having the largest impact on flowering time (Quinby and Karper 1945; Quinby 1966; Quinby 1967). Two additional maturity loci were discovered by Rooney and Aydin (1999) which further lengthen vegetative growth and were designated *Ma5* and *Ma6*. The *Ma5* and *Ma6* maturity genes, unlike the previously discovered maturity loci, function as an allelic complex in which at least one dominant allele is required at each loci to induce the photoperiod response (Rooney et al., 1999; Murphy et al., 2011). The discovery and implementation of these and possibly additional maturity loci is critical in maximizing biomass production in energy sorghums.

Photoperiod responses to day length and their effect on plant phenotype however are not restricted solely to sorghum. Orthologs of genes controlling photoperiod responses in sorghum have been found in other grasses such as *Arabidopsis*, rice and maize (Murphy et al., 2011). In addition to Murphy's (et al., 2011) findings, previous comparative mapping between sorghum, rice and maize revealed correspondence of height and maturity QTL (Lin et al., 1995). Furthermore, similar flowering time QTL between sorghum and sugarcane have also been identified (Ming et al., 2002). This indicates that maturity effects on plant phenotype observed within other Poaceae species may be relevant in explanation of phenotypic maturity responses in sorghum and allow for the transferability of molecular tools across species with co-linearity (Feltus et al., 2006).

The objectives of this study were to complete a QTL analysis for phenotypic traits influencing biomass yield in a population consisting of photoperiod sensitive recombinant inbred lines which do not segregate for maturity under long-day conditions. In addition to

evaluation under long days, the population was also evaluated in short and transitional day length environments to determine the relative effect of day length on plant phenotype and influence on detection of QTL for yield traits.

Materials and Methods

Plant Material and Experimental Design

R07018 and R07020 were selected as parents for the RIL mapping population. R07018 is white seeded, tan plant Guinea type sorghum while R07020 is a red seeded, purple plant Caudatum type sorghum. Under long days, both parents are tall, photoperiod sensitive (PS) and accumulate high biomass yields. Under short days both genotypes flower early, but R07018 is slightly taller and later flowering than R07020. The initial hybrid of these lines was produced under short days in Puerto Rico; all subsequent generations were advanced via head to row random selection in winter nurseries in Guayanilla, PR. A total of 100 $F_{2:5}$ RILs were generated.

The RIL population and both parental lines were planted in three environments. The first evaluation was a long-day environment in College Station, TX that was planted on April 12, 2011. The second environment was a transitional day length environment (from long to short days) in Weslaco, TX that was planted on August 11, 2011. During the growing season in this environment, day length reduced from ~13h at planting to ~11h at harvest and this induced reproductive growth at varying times within the population. Finally, a true short-day environment was planted in Guayanilla, PR on December 22, 2011. In this season, day lengths were less than 12 hours through most of the production season; therefore timing of

transition to reproductive growth was not influenced by the long-day sensitivity of these genotypes.

In all locations, the field plot design was a randomized complete block with two replications. In College Station and Weslaco, experimental units (plots) were 6.1m long with spacing of 0.85m between rows. In Puerto Rico each plot was 4.4m long with 0.85m between rows. The soil type in the College Station, Weslaco and Puerto Rico trials was Ships Clay Loam, Raymondville Clay Loam, and Constancia Silty Clay, respectively. Agronomic practices followed standard practices for each environment and included supplemental irrigation as needed to maximize productivity. Irrigation in both Texas environments was applied via flood irrigation while irrigation in Puerto Rico was via drip tape.

Measurement of Phenotypic Traits

Agronomic notes taken prior to harvest included; stem diameter, internode length, plant height, panicle exertion, and days to flower. Panicle exertion and days to flower were only taken in the Weslaco and Puerto Rico environments as the parental lines and population did not flower in College Station. Stem diameter and internode length were measured using a digital caliper and ruler, respectively, at the third internode from the base of the plant and recorded as an average of three stalks from each plot. Plant height was recorded as the average height of each plot from the base of the plant to the top of the panicle, except in College Station where plant height was recorded from the base of the plant to the bottom of the vegetative whorl due to the lack of panicle initiation. Panicle exertion was measured as the average peduncle length from the flag leaf to the lowest rachis branch of the panicle in

each plot. Flowering time was recorded as the day where 50% of the panicles within a plot were at mid-anthesis following planting.

Plots in the College Station trial were harvested 126 d after planting as photoperiod sensitive plants in this location transitioned from the log to lag phase of growth (Hoffmann, 2012). Trials in Weslaco were harvested 111 d after planting, approximately 30 d after the last entry flowered. In Puerto Rico plots were harvested 112 days after planting and coincided with grain sorghum reaching full maturity.

At each location, a 2m length of each plot was harvested for total biomass yield, juice yield and brix. In all environments total biomass yield included all harvested aerial plant material. Following measurement of total biomass yield, leaves were stripped and the stalks were weighed again and recorded as stalk yield. Leaf yield was determined by subtracting stalk yield from recorded total biomass yield. For environments in which panicles were present, panicles were removed and their yield was determined by subtracting their weight from total biomass yield. The stalks were then crushed to express juice using a three roller mill. In College Station and Weslaco, stalks were crushed using an Ampro Sugarcane Crusher Diamond Model (Ampro Exports; New Delhi, India). Crushing in Puerto Rico was performed using a food grade table-top sugarcane press. In College Station and Weslaco, juice volume from each plot was recorded in milliliters (ml) using a graduated cylinder. For all environments brix content (percent soluble solids in the collected juice) was measured with a PAL-1 digital pocket refractometer (ATAGO Co., LTD; Itabashi-ku, Japan). After crushing, bagasse samples were collected and weighed in grams (g) then dried in a Grieve model SC-400 forced convection dryer (The Grieve Corporation, Round Lake, IL) at 50-57°C for four days. Dried samples were re-weighed and using fresh and dry sample weights,

moisture content of the bagasse was calculated. In Puerto Rico, stalk yield and juice volume were not recorded.

Statistical Analysis

To account for spatial variation within field locations in all environments, nearest neighbor analysis was performed using Agrobase Generation II statistical analysis software (Agronomix Software Inc., Winnipeg, MB, Canada) followed by analysis of variance. Nearest neighbor analysis reduces variation within and between replications possibly due to environmental factors that are not uniformly present throughout the trial, while analysis of variance within and across environments allows partitioning of factors effecting observed variance to determine the relative effect each of these factors has on phenotypic measurements. Using mean squares from analysis of variance, variance components of each trait were estimated within and across locations. Variance of traits within locations was estimated as:

$$\sigma^2_{Trait} = \sigma^2_G + \sigma^2_{error}$$

Variance of traits across location was estimated as:

$$\sigma^2_{Trait} = \sigma^2_G + \sigma^2_E + \sigma^2_{GxE} + \sigma^2_{error}$$

where σ^2represents the variance due to genotype, σ^2_E is the variance due to the environment (location), σ^2_{GxE} is the variance due to the genotype by environment interaction and σ^2_{error} is the variance due to experimental error. Because replications were a non-significant source of variation (P=0.05) within each environment they were not included in the estimation of

variance components within locations. Using the variance estimates, the percent variation explained by each factor was determined as:

Percent variance within locations = $\sigma^2_i / \sum \sigma^2_{i +} \sigma^2_{error}$

Percent variance across locations = $\sigma^2_i / \sum \sigma^2_{i +} \sigma^2_{j +} \sigma^2_{k+} \sigma^2_{error}$

Where σ^2_i represents the variance estimate of the component being determined and $\sum (\sigma^2_{X)ijk +} \sigma^2_{error}$ is the sum of all variance estimates.

Broad sense heritability estimates (H^2) were calculated to determine how much of the observed variance was attributable to genetic variance. Broad sense heritability estimates were determined both within and across environments and combined heritability estimates were calculated for traits evaluated in all three environments. All heritability estimates were calculated using Agrobase Generation II as the proportion of total variance due to genetic differences. Within environments these estimates were calculated as:

$H^2 = \sigma^2_{G /} [\sigma^2_G + (\sigma^2_{error/R)}]$

Where σ^2_G is the variance due to genotype, σ^2_{error} is the variance due to experimental and/or residual error and R is the number of replications within each environment. Combined analysis of heritability included the variance of genetic × environment interaction and was calculated as:

$H^2 = \sigma^2_{G /} [\sigma^2_G + (\sigma^2_{GxE/E}) + (\sigma^2_{error/ER)}]$

Where σ^2_G is the variance due to genotype, σ^2_{error} is the variance due to experimental and/or residual error, E is the number of environments and R is the number of replications within each environment.

Population and entry means for all traits were calculated using JMP 9.0 (SAS Institute Inc.; Cary, NC) to determine population performance relevant to each parent (Table 1). Entry means for each environment were used in QTL analysis.

DNA Extraction and Map Construction

High-quality DNA was isolated from fresh leaf tissue of parental lines and RILs utilizing the FastDNA SPIN Kit (MP Biomedicals, Santa Ana, CA. Eight to 10 seeds of each parent or RIL were grown in petri dishes on moistened paper towels for 4 days prior to tissue harvest. Tissue from a minimum of 6 seedlings was collected and pooled for DNA extraction. Restriction-site-associated DNA template libraries were prepared from extracted DNA using the method of Morishige et al. (2013) for sequencing on an Illumina® GAIIx (Illumina, San Diego, CA). Using Illumina's Pipeline V1.5 sequence analysis software, text files resulting from genomic sequencing were produced containing 76 bp sequences which were then processed using a series of custom python scripts. Unique 4 bp identification tags were used to sort individual sequences based on the parental line or progeny from which they originated. The 4 bp identification tag used to associate sequences to individual genotypes was removed resulting in 72 bp sequences specific to individual genotypes. Sequences obtained more than 3 times from an individual genotype were aligned to the BTx623 genome sequence by BLASTN analysis (P. Klein, personal communication). Aligned sequences were then manually examined to identify location of their alignment to unique positions within the genome. Custom Perl scripts were used to identify polymorphic sequences between R07018 and R07020 through pairwise comparison of parental sequences to those aligned to BTx623 (P. Klein, personal communication). These polymorphisms were used to identify which alleles in individual RILs correspond to each parental line (Bioinformatic

analysis and polymorphism discovery were conducted by Patricia Klein, Ph.D.). Genetic map construction was performed using the resulting data and is detailed in the results section of this chapter.

QTL Analysis

Phenotypic data (adjusted trait entry means from each location) and genetic map data were imported to Windows QTL cartographer (Wang, et al., 2007) to identify significant QTL. Composite interval mapping (CIM) was used to identify significant QTL within each location as it is the most robust means of QTL detection through the reduction of residual variance not accounted for in other analyses (Haley and Knott, 1992). Significance thresholds for each QTL were determined through 1,000 permutations of the data at an alpha confidence level of 0.05 (Churchill and Doerge, 1994). Controls set within CIM analysis consisted of forward and backward regression using 5 control markers with 'into' and 'out' probabilities of 0.05, a window size of 10 cM with a walk speed of 1 cM. Quantitative trait loci with a LOD equaling or surpassing the threshold for each trait were reported as significant. A measurement of two LOD units away from the QTL peak (LOD 2 QTL interval) was used for reporting QTL size and R^2 was reported to determine the percent variation explained by each QTL for each trait. The additive effect of each QTL on plant phenotype was reported for each trait. Analysis of this population as an RIL prevented detection of dominance effects.

Results and Discussion

Phenotypic Analysis

Brix, internode length, and stem diameter were the only traits in which parental lines exhibited consistent behavior in reference to each other across environments (Table 1). In the long-day environment of College Station R07020 was taller than R07018 yet the opposite was true in the transitional and short-day environments of Weslaco and Puerto Rico. This was consistent with previous phenotyping by Rooney (personal communication) in which R.07018 was slightly taller than R.07020 under short-day growing conditions.

In Weslaco, R07020 flowered later than R07018 (Table 1). R.07018 had previously been phenotyped as later maturing than R07020 (Rooney, personal communication) which was consistent with flowering times recorded in Puerto Rico in this study. Flowering time of the RIL population was reduced (27 days for population mean) in the short-day environment when compared to the transitional day lengths of Weslaco (Table 1). Parental performance for the remaining traits could not be attributed specifically to the selection history or day length of each environment. Some traits indicated one parent to be the higher expressing parent in the long and short day environments while being the lower of the two in the transitional environment of Weslaco.

It was observed that the population means for some traits exceeded or failed to reach values recorded for either parent. With the exception of plant height, panicle yield, biomass moisture percentage and juice yield, traits exhibiting population means above or below parental values varied between environments (Table 1). Recorded means above or below the parental values may be due to a large number of individuals within the population having

higher recorded values than either parent. The proportion to which mean values of certain traits recorded for the population exceeded those of the parental lines was not consistent across all environments. This indicates that certain individuals within the population are better suited to specific environments which allowed an increased expression of those traits within the individuals. The minimum and maximum values recorded for all traits in the RIL population exceeded those observed for both parents, indicating transgressive segregation of the population.

These shifts in parental and RIL performance across environments indicate that not only individual genotype but also varying day-lengths and/or other environmental conditions directly influenced plant phenotype. Which factors most directly influence moisture content are more difficult to discern as handling of materials between harvest and processing and the length of time between the two can greatly effect these measurements. Traits influenced by plant moisture, i.e. volume and bagasse moisture percentage, are more difficult to compare across environments due to differences in timing between harvesting and crushing at each location.

Analysis of Variance Components

Within environments genotypes were a significant source of variation for all traits, and replication effects were not a significant source of variation (Table 2). Traits associated with plant moisture (i.e. bagasse moisture and volume) had a lower proportion of the variation associated with genotype (Table 2), presumably due to variation in time between harvests of individuals and processing.

In combined analysis, genetic effects were significant for exertion, bagasse moisture, brix and internode length (Table 2). However, environment and genotype × environment effects account for the majority of the variance for every trait (Table 2). Environment and genotype × environment effects were significant for all measured traits and for many traits environmental factors accounted for over 90% of the observed variance. Compared to effects measured within locations, residual error across locations was greatly reduced. The drastic differences between these three environments and their significant effect on plant phenotype combined with the significant effect of genotypes within each environment increased the amount of observed variance due to these factors. This reduces the relative effect of genotype on observed phenotypic variance across environments and renders them for most traits as a non-significant source of variation.

With such a large amount of phenotypic variance due to genotypic × environmental effects it is concluded that comparison of genotypes across locations is not appropriate within this population. The overwhelming effect of environment and genotype × environment interaction make combined analysis of genotypes across environments suspect.

Heritability Estimates

Heritability estimates ranged from 0.00 to 1.00 across all traits and environments (Table 3). For some traits, heritability estimates were quite different between environments (Table 3). For traits such as height, it is well known that long-day environments will increase biomass production due to delayed flowering (Burks et al., 2013). Thus, it was not surprising in College Station to have a low heritability estimate (0.39) for height as this is a long day environment. The high heritability of flowering time in Puerto Rico (0.65) was also expected

due to the reduction in photoperiod effect on maturity thereby increasing genetic expression of flowering across the population. Some traits had higher heritability estimates in the transitional environment than either the long or short-day growing conditions. For example, high heritability estimates for days to flower, internode length and stem diameter in the transitional day length environment indicate that alteration of day lengths during the growing season from above to below the photoperiod threshold allows a higher rate of genetic expression in the phenotype. This is to say that somewhat consistent day lengths throughout the growing season decrease genetic variance of some traits and not others. This could be the result of the effect of maturity on the expression of these traits.

Heritability estimates across all environments were lower for most traits (Table 3). This is consistent with what was expected as some traits were either expressed or suppressed between environments therefore reducing the genetic expression of such traits. The exception to this was internode length in which the combined estimate of heritability (0.69) was higher than any individual environment (Table 3).

Correlation of Traits

Only a few trait correlations were observed in multiple locations (Tables 4, 5, & 6). Positive correlations of biomass yield with stalk and leaf yield were significant in both College Station and Weslaco (yield components were not partitioned in Puerto Rico) (Tables 4 & 5). A positive correlation between juice volume and both biomass and stalk yield was also observed in both College Station and Weslaco. This is consistent with observations made by Murray et al. (2008). In both Weslaco and Puerto Rico, panicle yield was positively correlated with biomass yield (Table 6). This is consistent with observations made by

Murray et al. (2008) and Felderhoff et al. (2012) in which panicle yield was positively correlated with stalk yield. In Weslaco, panicle yield was positively correlated with stalk yield as well and would most likely have been in Puerto Rico had stalk yield been recorded. These associations of biomass and stalk yield were likely influenced by bird damage in Weslaco. Since early maturing seed concentrated birds, later flowering material exhibited less damage. This would allow for higher panicle weights for those individuals with extended vegetative growing periods.

Many traits were found to only be correlated in a single environment (Tables 4, 5, & 6). This was consistent with assumptions drawn from variance component analysis of strong environmental influences on plant phenotype. In College Station, brix was positively correlated with height, volume, biomass yield, and leaf yield while the only other correlation to brix was observed in Puerto Rico in which it was positively correlated with stem diameter. All brix correlations found in College Station were consistent with those found by Murray et al. (2008) and Felderhoff et al. (2012).

In College Station stem diameter was positively correlated with all vegetative yield components similar to reports by Murray et al. (2008) (Table 4). Negative correlations of exertion with leaf yield, panicle yield and internode length observed in Weslaco may be caused by associations with day length and/or subsequent biotic stress (i.e. bird damage) (Table 5). Exertion occurs only in reproductive growth which means that vegetative growth and accumulation of stalk biomass ceases. It occurred quite gradually in Weslaco (17 days) and this delay allowed those individuals flowering later to accumulate more leaf matter and internode elongation. This also caused earlier flowering RILs to produce mature seed earlier

which was eaten first by heavy bird infestation, and reduced recorded panicle yields at harvest.

Explanation of non-significant correlations to flowering time for most yield traits in Weslaco and Puerto Rico may be a function of values entered for days to flower (Tables 5 & 6). Individuals which flowered later had higher quantitative values for days to flower. A higher recorded value for flowering time (late flowering) will coincide with a higher recorded leaf yield due to prolonged vegetative growth.

In Puerto Rico, stem diameter was positively associated with higher brix which was consistent with findings by Murray et al. (2008) (Table 6). The absence of this association in Weslaco or College Station was surprising (Tables 4 & 5). Positive associations of internode length to height and negative associations of internode length to panicle yield in Puerto Rico appear to be a maturity effect, or lack thereof, on the expression of these traits (Table 6). The short-day environment of Puerto Rico allowed greater genetic expression of internode elongation within individuals than College Station and Weslaco. In Puerto Rico maturity was not suppressed or gradually initiated and thus the range of observed heights across the population was decreased although internodes were most elongated in this environment (Table 1). Due to the strong effect of environment on plant phenotype, combined correlations across environments were not performed.

Genetic Mapping

For genetic mapping 567 markers were scored through 97 $F_{4:5}$ individuals which resulted in a map consisting of 12 linkage groups spanning 1458 cM with an average of 2.63 cM between markers (Figure 1). A total of 16.5% of the loci were heterozygous slightly

higher than the theoretical 12.5% heterozygosity that would be expected for an F_4 population. The distribution of alleles from each parent was similar (38.2 % for R7018 and 42% for R07020). Roughly 3 percent of all marker data was reported as missing and not included in map construction.

QTL Mapping and Analysis

A total of 12 phenotypic QTL were identified across all environments (Table 7). Of these twelve, no co-localization was observed for QTL between environments although some QTL did show similar genetic locations within their respective environments. Seven QTL were detected in Weslaco, three in College Station and two in Puerto Rico. Of the traits for which QTL were identified, internode length and days to flower were the only traits in which QTL were identified in multiple environments.

For internode length, two QTL were identified in separate environments. One was found in College Station on chromosome 9 and one in Puerto Rico on chromosome 7 (Table 7). While both QTL explain roughly the same amount of variance for the trait (13 and 14%) in College Station R07020 alleles increased internode length while R07018 alleles increased internode length in Puerto Rico. Genetic explanation of this inverse phenotypic expression cannot be explained through this study.

A single QTL for days to flower was identified in both Weslaco and Puerto Rico (Table 7). In Weslaco this QTL was found on chromosome 2 and in Puerto Rico on chromosome 3. R07020 contributed the allele increasing flowering time in Weslaco while R07018 contributed to increased flowering time in Puerto Rico. As noted earlier, higher values recorded for days to flower indicated later flowering plants in this study. Alleles

responsible for later flowering in Puerto Rico were contributed by R07018. In Weslaco, however, R07020 was observed to be responsible for later flowering.

Two QTL identified in Weslaco for panicle weight mapped to near identical locations on chromosome 2 (Table 7). The likelihood of odds interval indicates the centimorgan distances for which the QTL span and these QTL overlap at 74.6cM. This overlap of QTL intervals indicates that this may be a single major effect QTL. As values which were recorded for panicle weights in Weslaco may have been adversely affected due to bird damage, variances in data may have caused this single QTL to appear as separate loci.

In Weslaco, single QTL for height and leaf weight were identified at the same peak centimorgan location on chromosome 10 (Table 7). R07020 contributed the allele for increased plant height and R07018 contributed the allele for increased leaf weight. While these two loci are located at the same region, no correlation was observed for these two traits in Weslaco (Table 5).

In College Station, a single QTL for stem diameter was identified on chromosome 3 and another QTL was detected for volume on chromosome 2 (Table 7). An allele contributed by R07020 increased stem diameter while an R07018 allele increased juice volume.

Single QTL for brix and stalk weight were observed in Weslaco; this was the only environment in which QTL for these traits were identified (Table 7). Because the majority of the QTL were detected in this environment, it implies that the decreasing day lengths in Weslaco facilitated differences among genotypes for many more agronomic traits than College Station or Puerto Rico. Conversely, the minimal differences in maturity make it

difficult to detect QTL for many of these traits, further confirming the influence of maturity on phenotypic traits.

Co-localization of QTL across studies is an indicator of QTL consistency although the lack of co-localization does not imply an invalid QTL. From this study, the brix QTL found in Weslaco on chromosome 3 (Table 7) was mapped to the same chromosome in studies by both Murray et al. (2008) and Felderhoff et al. (2012).

Conclusion

This study utilized a population which did not segregate for major loci controlling height and photoperiod sensitivity when grown under long days. The lack of segregation at these loci reduced the ability to detect major effect loci for both these traits and associated traits (i.e. biomass yield). This is caused by the genotypes in the study failing to enter reproductive growth when grown under long days. This in turn prevents differentiation of height and maturity between genetically different individuals.

Similar to long day growing conditions, phenotypic differences within this population when grown under short days also decreased our ability to discern between differing genotypes. While physiological maturity wass reached when grown under short days, the genetic influence on maturity wass reduced and with that our ability to detect it. Despite the reduced variation in plant phenotype in the short day environment of Puerto Rico, we were able to detect a single QTL for days to flower. While the short days reduced the overall number of QTL detected, flowering time is a major trait of interest to breeders and sorghum researchers alike. This ability to identify important QTL which control many traits is instrumental in continuing to understand the far reaching effects of maturity on plant

phenotype. Although the flowering time QTL detected in Puerto Rico has not been confirmed as a novel maturity loci it should not be assumed that additional maturity genes do not exist. Studies such as this one, which provide new methods for testing sorghum genotypes not previously researched, increase our ability to discover and analyze a broader range of genetic variation not previously available.

It is evident from this study that phenotypic variation resulting from the decreasing day lengths in Weslaco improved our ability to detect these QTL. Genetic diversity within the population is present however in the long and short day environments and is confirmed by our identification of QTL in Weslaco. This indicates that in the absence of phenotypic variance, the genetic influence on phenotype cannot be fully determined as it is masked by a lack of phenotypic variation. This does not however indicate that photoperiod insensitive populations must be used to evaluate these phenotypic traits. What is more important is determining the correct environment to express and detect phenotypic and genetic variation in the population. By reducing the phenotypic variation of the population our power to correlate genetic differences to phenotypic differences is reduced and thereby decreases the power and feasibility of mapping QTL in photoperiod sensitive populations in long day environments.

Utilizing photoperiod sensitive populations for trait loci identification increases our ability to identify traits in sorghum populations which have not been subjected to selection for insensitivity and will allow a broader testing of existing sorghum genotypes. While the number of QTL identified will decrease using photoperiod sensitive populations as observed in this study, the ability to find them does exist. By analyzing more photoperiod sensitive populations in environments of varying day length or maximizing the genetic diversity

present in photoperiod sensitive mapping populations, the number of loci identified will increase and this will provide a high level of validation for QTL if they are present in multiple populations. The use of long and short day environments should not, however, be excluded from such analysis as it provides a means of trait identification without the effect of maturity on plant phenotype.

CHAPTER III

IDENTIFICATION OF QTLS INFLUENCING BIOMASS COMPOSITION IN A PHOTOPERIOD SENSITIVE RIL SORGHUM POPULATION

Introduction

Energy sorghums are roughly divided into two types; sweet sorghum which contains high levels of soluble sugars and high biomass sorghums which are primarily ligno-cellulosic biomass (Rooney et al., 2007). Sweet sorghums were first used in the U.S. as a source of sweetener in the form of syrup. A limited amount of breeding to improve these sweet sorghums was completed in the mid-20th century. The biofuel industry is interested in sweet sorghum as a sugar source for ethanol production and while the syrup varieties are a logical starting point for improvement, the selection criteria for industrial sweet sorghums will be quite different (Rooney 2007). Sugars from industrial sweet or biomass sorghum genotypes must be high yielding, easily extracted and readily fermentable.

Much of the effort to improve bioenergy sorghum has focused on increasing biomass and juice yield. Studies to elucidate the genetic control of biomass quality and composition are more limited. However, because biomass conversion efficiency is influenced by biomass composition, it is important to assess the relative variation and genetic basis of sorghum composition (Monti et al., 2008). Sorghum biomass can accumulate significant quantities of both structural and non-structural carbohydrates; the exact composition of which is contingent on the type of sorghum that is being produced, its maturity at harvest and the environment in which it is grown. Comparison of biomass composition for nonstructural

carbohydrates indicates that sweet sorghums accumulate up to 25% more sugars than grain sorghums in the stalks (Murray et al., 2008). Murray et al. (2008) also reported that grain-type sorghums generally produce higher concentrations of lignin, cellulose and protein while sweet sorghums tend to have higher concentrations of hemicellulose. Compared to sweet sorghum, it is logical that biomass sorghums should have higher levels of structural carbohydrates due to lower sugar concentrations in the stalk. Breeding high biomass sorghums for increased yield and reduced lodging may result in levels of structural carbohydrates even higher than those of grain sorghums but these comparisons have not been tested.

The primary component of biomass is lignin and the structural carbohydrates (cellulose and hemicellulose). When separated from lignin, both cellulose and hemicellulose can be used to produce ethanol or other energy compounds. The simple linear structure of cellulose requires a few enzymes to break the glucose chain into individual molecules (Perez and Munoz-Dorado 2002). Hemicellulose however is more complex; it contains both xylan and glucan, with xylan being the most abundant sugar in hemicelluloses of herbaceous plants (Perez and Munoz-Dorado 2002). Its complex structure requires more enzymes for hydrolysis than cellulose.

To access both cellulose and hemicellulose, it must be separated from lignin. Lignin is an organic biopolymer molecule that provides strength and hydrophobic properties to the cell wall through cross-linking with cellulose and hemicelluloses (Theander et.al, 1993). This linkage concomitantly inhibits degradation by limiting the ability of enzymes to contact and degrade cellulose and hemicelluloses. While this resistance is necessary for plant survival, it certainly reduces the efficiency of the conversion process (Corn 2009).

While carbohydrates and lignin are the primary components for energy production, several other compounds must be considered in bioenergy sorghums as they affect processing or sustainability. Proteins are linear polymers consisting of amino acids and play an important role in cell structure and function. Elevated levels of protein present in biomass reduce cellulose digestibility as well as fermentation efficiency during ethanol production making it undesirable in a biomass feedstock (Murray et al., 2008).

Ash content within biofuel feedstocks is problematic for processors during combustion of biomass because increased ash content negatively affects heating value; every 1% increase in ash concentration decreases heating value by $0.2MJkg^{-1}$ (Monti et al., 2008). Other common problems associated with ash include mineral deposits and corrosion of metal surfaces of processing equipment. The highest levels of ash are consistently found in the leaves of six of the major energy crops including sorghum (Monti et la., 2008; Olson et al., 2012). Partitioned between the leaves, stems and reproductive organs, the ash content present in leaf matter was almost double that of the stems and roughly 50% greater than the reproductive organs. Olson et al. (2012) reported that ash content was twice as high in the leaves as in the stalk of sorghum. Given that nitrogen content is higher in leaves as well (Olson et al., 2012), it is logical to return as much foliage as possible to the field.

Genetic and environmental factors are known to influence plant composition (Rooney et al., 2007; Monti et al., 2008; Corn 2009). Variation in biomass composition of forage sorghum has been studied quite extensively and recently described for 22 commercially available forage sorghum hybrids (Dahlberg et al., 2011). While this study did identify genotypic variation it sampled only a small portion of the sorghum genetic diversity. Stefaniak et al. (2012) however observed high levels of compositional variance when testing

108 diverse sorghum genotypes which were due to genotypic, environmental and genotype × environment interactions. Although factors other than genotype, such as environment, influence plant composition the relative contribution of these factors in regards to bioenergy sorghum have not been determined. The relative magnitude these effects have on plant composition is critical to determine our ability to alter biomass composition (Corn 2009).

To date, QTL mapping efforts for bioenergy sorghum have utilized populations that segregate for both height and/or maturity. Given that these traits are pleiotropic for both biomass productivity and composition, they exert a large effect and may mask additional QTL for biomass composition (Ritter et al., 2008; Murray et al., 2008; Felderhoff et al., 2012). Thus, there is a need to identify QTL influencing biomass composition and determine the relative effect of factors contributing to their variance independent of the effects of maturity and plant height.

The objectives of this study were to detect QTL for biomass composition in a population of photoperiod sensitive recombinant inbred lines which do not segregate for maturity under long-day conditions. In addition to evaluation under long days, the population was evaluated in short and transitional day length environments to determine day length and environmental effects on plant composition and QTL detection.

Materials and Methods

Plant Material and Experimental Design

The parental lines and RIL progeny used for this study were the same as those used to map QTL for phenotypic traits in Chapter II. The experimental design was as described in Chapter II.

Measurement of Composition Traits

Plots in three environments were harvested as described in Chapter II.

To measure plant composition, dried bagasse samples were ground using a Wiley knife mill (Thomas Scientific Swedesboro, NJ) to pass through a 2mm sieve. Biomass composition was estimated through near infrared spectroscopy (NIR) in which dried, ground samples were scanned using a Foss XDS (Foss Hillered, Denmark) with ISI-scan software that measured reflectance at wavelengths between 400-2500nm. Predictions for biomass composition were based on a calibration curve developed through collaborative research between Texas A&M University and the National Renewable Energy Laboratory (Wolfrum et al., 2013). While additional components were measured, composition analysis on a percent dry weight basis is provided for ash, protein, sucrose, lignin, glucan, and xylan. Because not all of the components estimated are presented herein, the composition percentages do not total 100%.

Statistical Analysis

Statistical analysis on composition traits was conducted as detailed in Chapter II for phenotypic traits. Entry means for each environment were used in compositional QTL analysis.

QTL Analysis of Composition Traits

Composition data and genetic linkage map data were imported to Windows QTL cartographer to identify significant QTL using spatially adjusted trait entry means from each environment. Composite interval mapping (CIM) was used to identify significant QTL within each location as described in Chapter II.

Results and Discussion

Phenotypic Analysis

Composition varied across environments for the parents and RILs. Ash content in R07020 was lower in Weslaco and Puerto Rico than in College Station. In R07018 as day lengths were reduced from College Station to Puerto Rico ash content increased (Table 8). In general, lignin, xylan and glucan content increased in both parental lines as day lengths were reduced (Table 8); the exception was the glucan content in R07018 in Weslaco which dropped slightly. Protein levels in parental lines appear similar in College Station but were highest in R07018 in Weslaco and Puerto Rico. The opposite was true for sucrose in which the highest values for R07020 were observed in Weslaco while R07018 expressed its highest sucrose values in College Station. The lowest concentrations of protein and sucrose for both parental lines were in Puerto Rico, likely due to the production of grain in this environment.

Ash, lignin, xylan, and glucan contents were highest in Puerto Rico for the RIL population (Table 8). Relative to the other locales, the ash content was somewhat higher in Puerto Rico because leaves were not removed prior to crushing or sample collection. Leaves have almost twice the ash content of stems (Monti et al., 2008). With regard to lignin, xylan, and glucan it is logical to conclude that the reduced mean biomass yield and stem diameter

recorded in Puerto Rico (Table 1) would result in elevated percentages of these components due to more plant material consisting of structural carbohydrates.

For the RIL population, protein content was highest in College Station while the highest observed sucrose level was recorded in Weslaco (Table 8). The population ranges recorded for non-structural carbohydrates overlap for all environments indicating relatively large variances for these components within environments but reduced variance of composition percentages between environments. At least one RIL was equal to or exceeded the minimum and maximum values of both parental lines for all recorded compositional components in all environments, indicating bimodal transgressive segregation of the population. Presence of such segregation indicates genetic variability for compositional traits between the parental lines.

Analysis of Variance Components

For all compositional components genotype was a significant effect in all environments (Table 9). In College Station and Weslaco, genotypes accounted for over 60% of all observed compositional variance and over 70% in Puerto Rico. This was not surprising given that plant phenotype directly influences plant composition (Murray et al., 2008; Felderhoff et al., 2012; Stefaniak et al., 2012).

In the combined analysis of compositional components, the majority of variation was associated with the environment (Table 9), which has been previously observed in composition studies of plant biomass (Murray et al., 2008). However, both genotype and the genotype × environment interaction were also significant for all compositional traits except protein. One would assume that inferences made on compositional data from combined

analysis would be reliable due to significance of genetic effects being maintained from individual environments. However, the presence of highly significant genotype × environment interactions for all compositional components (Table 9) casts doubt on any inferences drawn from combined analysis. This is especially relevant to populations similar to this one, in which differing environments mask or prevent expression of genotypic responses influencing biomass yield and maturity.

Heritability Estimates

Heritability was moderate to high within each environment for most of the compositional traits, ranging from a low of 0.45 for glucan to a high of 0.68 for ash in College Station. By location, the highest average heritability was in Puerto Rico and the lowest was in College Station (Table 10). In the combined analysis, all heritability estimates dropped due the significant genotype × environment interaction.

For sucrose, glucan and xylan, heritability increased as day lengths within growing environments decreased (Table 10). As maturity was prevented in the long day environment and timing of flowering was relatively synchronous in the short day environment, this would imply that genetic expression of sucrose, glucan, and xylan are dependent on timing of flowering. This assumption however cannot be validated through currently published research and to our knowledge this is the first study to evaluate trait heritability in the presence and absence of plant maturity.

The heritability of protein content was relatively consistent throughout all environments while heritability of other compositional traits such as ash appeared more variable across environments (Table 10). The presence of leaf matter may have increased the

heritability of ash in Puerto Rico because leaves contain higher levels of ash than stalks (Monti et al., 2008).

Heritability for lignin ranged from 0.47 in Weslaco to 0.60 in Puerto Rico implying that reproductive growth increases heritability of lignin (Table 10). However, lignin heritability in Weslaco appears similar to College Station although no statistical separation was calculated. Had genetic expression been truly dependent on maturity, then the heritability of lignin in Weslaco should be between College Station and Puerto Rico. These results indicate that factors outside of maturity or in conjunction with it influence lignin heritability.

Trait Correlations

Sucrose levels were negatively associated with lignin, xylan, and glucan in all environments (Tables 11, 12 & 13). These findings are consistent with those observed by Stefaniak et al. (2012) although in that study, association between structural carbohydrates were partitioned by sorghum type but non- structural carbohydrate correlations, such as sucrose, were grouped across all sorghum types tested.

Ash was positively associated with protein content in College Station and Puerto Rico yet negatively associated with lignin in College Station (Tables 11 & 13). This indicates that flowering time and/or plant maturity directly impact proportions of these components in relation to each other. The positive associations of ash observed in Puerto Rico however may have been influenced by the leaves not being removed prior to sampling (Table 13). This elevated the overall levels of ash recorded in relation to protein in Puerto Rico. The

correlations of ash with protein and lignin were consistent with those found by Stefaniak et al. (2012).

In all environments protein was negatively correlated with sucrose (Tables 11, 12, & 13). This finding of negative correlation between protein and sucrose was similar to that observed by Stefaniak et al. (2012). These associations were not affected by growing environment day length nor were these interactions influenced by plant maturity or yield components affected by maturity.

Analysis of protein and the structural carbohydrates lignin, xylan, and glucan indicate that only glucan appears to be correlated to protein and is consistent across all environments regardless of day length (Tables 11, 12, & 13). Lignin and xylan however show no association with protein in any of the environments. Lastly, a negative association between ash and glucan was reported for College Station (Table 11) yet no such correlation was found in other environments. This negative interaction is impacted by suppression of plant maturity and/or yield components resulting from prolonged vegetative growth as this association was not observed in environments in which plants reached physiological maturity.

Differences of identified correlations in the study by Stefaniak et al. (2012) to those presented herein may be due to differences in analytical methods. Correlations identified by Stefaniak et al. (2012) when partitioned by sorghum type were consistent with our results yet their grouping of all sorghum types for non-structural carbohydrate associations reduced the co-linearity of the two studies.

QTL Mapping and Analysis

A total of 24 compositional QTL were detected across the three environments (Table 14); ten QTL were detected in both College Station and Puerto Rico with the remainder detected in Weslaco. At least one QTL was detected for every compositional component in at least one environment and QTL for ash, sucrose, and lignin were identified in all environments (Table 14). Most of these QTL account for approximately 10% of the variation associated with the trait, with a single QTL for protein accounting for 21% and a QTL for ash in College Station that accounted for 8% (Table 14). While there are likely many additional QTL with smaller effects, the statistical power of this study (population size and replications) are not sufficient to detect them.

There was some co-localization of QTL in this study. A QTL for ash on chromosome 6 was detected in both College Station and Puerto Rico (Table 14) with a peak at 85.4. In both situations, the R07018 allele increased ash content between 0.2 and 0.28%. The consistency of the QTL across environments indicates that this region of the genome is regularly affecting ash content in biomass sorghum and would be a candidate for marker assisted selection. A QTL for protein were identified in similar locations on chromosome 4 in College Station and Puerto Rico (Table 14). In College Station the QTL peak was at 37.1 cM and at 39.1 cM in Puerto Rico. This increase in protein was contributed by R07020 for both QTL although the relative effect varied. In College Station this QTL increased protein content 0.22% while in Puerto Rico the change was 0.13%.

The most QTL for any component were detected for ash content (Table 14). A total of 8 QTL were found across the three environments (Table 14). Four were identified in

College Station on chromosomes 4, 5 and 6. Three QTL for ash were found in Puerto Rico as well as a single ash QTL in Weslaco. Two of the ash QTL identified in College Station were contributed by R07020, increasing ash content by 0.24 and 0.29% while the other two were contributed by R07018 and increased ash by 0.20 and 0.2%. The three ash QTL found in Puerto Rico on chromosomes 6 & 7, contributed by R07018, increased ash content between 0.19 and 0.28% while the one detected in Weslaco on chromosome 2 and contributed by R07020 increased ash content by 0.19%. The overall number and distribution of QTL across the genome indicate that this component is controlled by many QTL with smaller effects as no QTL accounted for more than 16% of the total variance. Contribution of alleles increasing ash content by both parents indicate that these QTL are present in both parental lines and the segregation of these alleles contributed to the wide ranges of ash content observed in the RIL population.

Four sucrose QTL were identified across all environments (Table 14). Two QTL were in College Station on chromosomes 3 and 6. The locus identified on chromosome 3 in College Station was contributed by R07018 and increased sucrose content 0.56% and the one identified on chromosome 6 was contributed by R07020 and increased sucrose by 0.54%. Both QTL identified in College Station accounted for less than 10% of the variation in total sucrose content. The QTL identified in Puerto Rico, contributed by R07018, increased sucrose by 0.53% while the one identified in Weslaco was contributed by R07020 and increased sucrose content by 0.88%. While these QTL had the largest effect of any on component concentration, their relative contribution to variance of the population was modest; no QTL accounted for more than 13% of variance (Table 14).

Five lignin QTL were identified with at least one loci being detected in each environment (Table 14).The two lignin QTL identified in College Station were located on chromosomes 4 and 7 while the lignin QTL in Puerto Rico were identified on chromosomes 5 and 8. All loci increased lignin concentration and with the exception of the one observed in Weslaco on chromosome 1, were contributed by R07020. Increasing effects of these loci on lignin ranged from 0.24 to 0.35% with the highest being in Weslaco and contributed by R07018. None of the QTL identified explained more than 14% of total variance for lignin. The relatively little variance explained by these QTL was not surprising as concentrations and ranges of lignin content within the population and parental lines was minimal. Comparisons of this study to that by Murray et al. (2008) and Felderhoff et al. (2012) indicate similar chromosomal locations of the lignin QTL on chromosome 7 in all studies. Furthermore, of the 10 lignin biosynthesis gene families identified in sorghum 7 are on chromosome 4, 5 are on chromosome 7, 2 are on chromosome 5 and 1 is on chromosome 8 (Zhanyou et al., 2009). Thus it can be confirmed that the chromosomes, on which lignin QTL were identified in all environments of this study, do contain known lignin genes.

Two QTL for xylan were identified in Puerto Rico while a single QTL for glucan was identified in both Puerto Rico and Weslaco (Table 14). QTL for xylan identified in Puerto Rico were contributed by R07020 with increasing effects of 0.16 and 0.17% and both alleles accounted for less than 0.1% of variation for the trait. R07020 contributed both alleles increasing glucan concentration and both loci accounted for 0.14% of the variance for the trait. Identification of QTL for both glucan and xylan indicate genotypic differences between the parental lines controlling the trait. The lack of QTL identified for these traits in College

Station indicate that while genetic differences for the components do exist, segregation of alleles controlling these traits within the RIL population is reduced in long day environments.

Conclusion

The present results confirm that it is possible to detect compositional QTL in populations which do not segregate for maturity. While compositional QTL were detected in this study, expression and/or suppression of maturity appeared to affect the ability to detect them. When analyzing phenotypic traits, the transitional environment allowed detection of more loci yet the short and long-day environments, in which flowering was prevented or more uniform, expressed more compositional QTL. It can be concluded from this data that photoperiod sensitive genotypes express small yet higher levels of genetic variance for compositional components than phenotypic traits when the effect of maturity is minimized in the population. Reduction of population structure in QTL studies of photoperiod sensitive individuals through the use of association panels will increase genetic diversity which is reduced by selection within a bi-parental population of similar phenotypes.

The GWAS study by Morris et al. (2013) indicates that strong population structure can and does exist within differing sorghum types. Selection for common agronomic traits across these sorghum types strengthens population structure and reduces genetic variation for those traits. The ability to identify QTL will be increased through the use of sorghum types similar in phenotype having distinct genetic backgrounds with reduced selection. The feasibility of QTL identification in photoperiod sensitive RIL populations can be validated through association studies aimed at identifying similar traits in more diverse photoperiod sensitive individuals. The QTL identified in this study can also be validated through similar

studies using photoperiod sensitive RIL populations which will confirm or reject the conclusion that reduction of maturity and genetic diversity reduce the ability to identify the genetic basis of agronomic traits.

CHAPTER IV

CHARACTERIZATION AND ANALYSIS OF *SORGHUM* × *SACCHARUM* F_1 INTERGENERIC HYBRIDS

Introduction

Increasing worldwide demand for food and fuel has placed renewed emphasis on creating cultivated crop species with higher yields and improved tolerances to biotic and abiotic stresses (Maqbool et al., 2001; Mathews et al., 2000; Dillon et al., 2007). Two species that have received growing attention to achieve this goal are sorghum and sugarcane (Ahn and Tanksley 1993; Giussani et al., 2001; Paterson 2008; Paterson et al., 2009).

Sorghum (*S. bicolor*) is the world's fifth most important grain crop based on production and is second as a source of U.S. grain-based ethanol (Paterson, 2008). Drought and heat stress tolerance in sorghum are valued in current production regions and ensure its continued production. Sugarcane (*Saccharum* spp. L.) however ranks first in global production as a sugar producing crop, accounting for 60% of raw sugar consumed worldwide, and is grown in more than ninety countries worldwide (Grivet and Arruda 2001; Dillon et al., 2007). A significant amount of sugarcane is now used for ethanol production, especially in Brazil (Goldemberg et al., 2004).

Introduction of traits improving plant fitness have in recent decades allowed sorghum to become adapted to a more diverse range of growing conditions. Increases in grain yield and improved tolerance to biotic and abiotic stresses have also been achieved during this improvement of environmental adaptation (Maqbool et al., 2001). Sugarcane production regions however have remained relatively unchanged despite breeding efforts to improve

crop performance and adaptability. Despite the economic importance of sugarcane, genomic research has yielded less improvement than that observed in other agronomically important crops, most likely due to its complex genome (Grivet and Arruda 2001.

Hybridization within a species has long been the basis for crop improvement (Kuhlman et al., 2008; Hodnett et al., 2010). While it is less frequent, interspecific and even intergeneric hybridization has been crucial to break through improvement barriers and have led to innovations in crops. For example, the crop Triticale (*Triticale hexaploide* Lart.) was the first man-made crop species that was developed through hybridization of wheat and rye (*Secale cereal* L.) (Zenkteler and Nitzsche 1984).

Intergeneric hybridization within Poaceae, specifically sorghum and sugarcane, may provide opportunities for trait introgression between genera and assist in improving the quality and utility of both crops (Hodnett et al., 2010). Successful hybridization between sorghum and sugarcane has previously been reported, but the frequency of hybrid production was minimal (Gupta et al., 1978), trait introgression between the species was never confirmed and the hybrids were of little breeding value (Nair 1999).

The development of Tx3361 (Kuhlman and Rooney, 2011), a *S. bicolor* genotype containing the mutant allele *iap* (Inhibition of Alien Pollen) facilitated the intergeneric hybridization of sorghum and sugarcane. Hodnett et al. (2010) demonstrated the utility of the *iap* mutant in Tx3361 by producing over 14,141 hybrid seed from 252 *Sorghum* × *Saccharum* crosses over a three year period. With a proven method to produce large numbers of hybrids, it is now possible to identify superior genotypes for both propagation and introgression.

Given the diversity within sugarcane, segregation within F_1 plants is expected and variation in performance among F_1s should be expected. Characterization and screening of the F_1 hybrids for desirable agronomic traits may provide opportunity for development of a new hybrid crop (Hodnett et al., 2010). Following selection of elite hybrids it is critical that these lines be evaluated in conjunction with sugarcane and sorghum to assess their relative value.

From the hybrids produce by Hodnett et al. (2010), the objectives of this study were to evaluate and select elite F_1 sorghum × sugarcane hybrids from observation plots and then complete replicated testing of these lines to determine their agronomic desirability and utility in comparison to elite sugarcane cultivars. The goal is to confirm genetic variation, phenotypic variation and to compare their performance with elite sugarcane varieties.

Materials and Methods

Plant Material

From the seed produced by Hodnett et al. (2010), a total of 493 confirmed F_1 *Sorghum* × *Saccharum* hybrids were transplanted from a greenhouse into a space plant nursery in College Station, TX in April of 2009. Row spacing was approximately 1.5m with plants spaced 2.4m apart within the row. The soil type in College Station was Ships Clay Loam and agronomic practices followed standard sorghum cultivation for this environment. This included supplemental irrigation as needed to maximize productivity which was applied through flood irrigation.

The nursery was allowed to grow through the summer and individual plant data was recorded in September 2009. Each plant in the nursery was evaluated based on general

agronomic appearance which included height, vigor, and general desirability. Plants were scored using a 1-9 scale (1 being most desirable). Of the 493 hybrids, the top six visually rated plants were selected for plant increase and further testing.

Six selections from the 2009 space plant nursery in College Station, TX; designated BSX0111, BSX3313, BSX4221, BSX5115, BSX7213 and BSX7413 were increased in a greenhouse in College Station, TX during the 2010 growing season. In November 2010 these seven genotypes were planted from cane in a randomized complete block design consisting of three replications in Weslaco, TX. Each experimental unit was 6.1m long with 1.5m between rows. The soil type in Weslaco was Raymondville Clay Loam and all plots were irrigated in conjunction with surrounding sugarcane through flood irrigation. Sugarcane accession TCP89-3505 is an elite cultivar developed by the Texas Agriculture Experiment Station Weslaco Center and registered in 2005 (Scott et al., 2005). It was planted in each of the three replications as a check for which to compare intergeneric hybrid growth and performance.

Data Observation

All entries in the study were harvested December 13, 2011 which is within the typical sugarcane harvest season in the Rio Grande Valley of Texas. Prior to harvest plant height, stem diameter, and internode length were recorded. Stem diameters and internode lengths were measured using a digital caliper and ruler, respectively, at the third internode from the base of the plant and recorded as an average of three readings from each plot. For entries that were still in vegetative growth, plant height was measured as the plot average from the

base of the plant to the base of the vegetative whorl. For the entry that flowered, plant height was measured from the base of the plant to the top of the panicle.

At harvest, one meter samples were clipped from each plot and weighed to determine total biomass yield. Leaves were then stripped and plots were re-weighed to determine stalk yield. Leaf yield was determined by subtracting stalk yield from the total biomass yield. Stalks were crushed to express juice using an Ampro Sugarcane Crusher Diamond Model (Ampro Exports; New Delhi, India). Total juice volume was measured and brix readings were recorded using a PAL-1 pocket refractometer (ATAGO Co., LTD; Itabashi-ku, Japan). Following crushing, bagasse samples were collected and weighed (in grams) and the sample was dried in a Grieve model SC-400 forced convection dryer (The Grieve Corporation, Round Lake, IL) at 50-57°C for a minimum of four days. Dried samples were weighed and moisture content of the bagasse was calculated based on the fresh and dry weights of the sample.

For composition analysis, dried bagasse samples were ground using a Wiley knife mill (Thomas Scientific Swedesboro, NJ) to pass through a 2mm sieve. Genotype composition was estimated through near infrared spectroscopy (NIR) in which dried, ground samples were scanned using a Foss XDS (Foss Hillered, Denmark) with ISI-scan software that measured reflectance at wavelengths between 400-2500nm. Predictions for biomass composition were based on a calibration curve developed through collaborative research between Texas A&M University and the National Renewable Energy Laboratory (Wolfrum et al., 2013). While additional components were measured, composition analysis on a percent dry weight basis is provided for ash, protein, sucrose, lignin, glucan, and xylan.

Because not all of the components estimated are presented herein, the composition percentages do not total 100%.

Statistical Analysis

Data was analyzed using JMP 9 statistical analysis software (SAS Institute Inc., Cary, NC). Means and standard deviations of measured traits were calculated for individual genotypes followed by analysis of variance. Trait mean squares resulting from the analysis of variance were used to estimate variance components. Trait variance was estimated as:

$$\sigma^2_{Trait} = \sigma^2_G + \sigma^2_R + \sigma^2_{error}$$

Where σ^2_G indicates variance due to genotype, σ^2_R is variance due to replication and σ^2_{error} indicates variance due to residual and/or experimental error. Variance components which lacked significance in analysis of variance were not included in the estimation of trait variance. Variance estimates were then used to estimate the percent variation explained by each component. Percent variation explained by each factor was determined as:

$$\text{Percent variation} = \sigma^2_i / \sum \sigma^2_i + \sigma^2_j + \sigma^2_{error}$$

Where σ^2_i is the variance estimate of the component being determined, σ^2_j is the variance estimate of the component not being calculated and σ^2_{error} is the variance estimate of experimental error. Explained variance was not reported for components which lacked significance in the analysis of variance. Repeatability of trait expression however can be calculated using similar methods to that of heritability estimation. Repeatability of traits was calculated as:

$$\text{Trait repeatability} = \sigma^2_{G} / [\sigma^2_G + (\sigma^2_{error/R})]$$

Where σ^2_G is the variance due to genotype, σ^2_{error} is the variance due to experimental and/or residual error and R is the number of replications in the test.

Pearson correlations were calculated to determine significance of interactions between measured agronomic traits. Measurements recorded for TCP89-3505 were not included in this analysis as to not confound interpretation of trait correlation for intergeneric hybrid genotypes. Inclusion of non-hybrid data in this analysis will decrease the ability to determine accurate trait correlations within the hybrids.

Results and Discussion

Phenotypic and Composition Analysis

Of the seven entries in the test, sugarcane cultivar TCP89-3505 produced the highest biomass yield, stalk yield, juice yield and sucrose concentrations and the lowest lignin concentrations of any entry (Table 15). While these are just a few of the desirable traits commonly found in elite sugarcane varieties adapted to the region they are among the most important for commercial sugarcane production (Berding 2004).

The number of stalks produced by BSX5115 was similar to that of the sugarcane check and its xylan content was the highest recorded for any entry in the test (Table 15). The percent glucan measured in BSX4221 was also similar to that of TCP89-3505. While these hybrids were not selected from the space plant nursery based on elevated levels of these traits it does indicate that for these traits the selected hybrids are comparable to the sugarcane entry.

For the remaining traits at least one hybrid was similar or superior to TCP89-3505 (Table 15). Based on the phenotypic and composition results it is possible to produce intergeneric hybrids between sorghum and sugarcane with desirable agronomic traits similar to that of commercially available sugarcane cultivars.

As this was the first comparison between sugarcane cultivars and intergeneric sorghum × sugarcane hybrids, expectations of overall F_1 hybrid performance was minimal. What was expected was that each F_1 hybrid would be comparable for at least one trait to the sugarcane check. While no individual outperformed the sugarcane check for all measured traits in this test most hybrids did exhibit one or more traits with similar or exceeding measurements to those recorded for the sugarcane check (Table 15). The exception to this was BSX3313 which was neither the highest nor lowest measured hybrid for any recorded trait

Analysis of Variance Components

Variance component analysis of individual traits is presented in Table 16. Indications from partitioning of each component are that for most traits the genotypic effect contributed significantly to the observed variance between entries. Height and leaf yield are the only traits in which replication contributed to observed differences. While significant, the percent variation explained by the effect of replication is less than 10% for both traits in which it was observed. The replication effect on plant height may be caused by the adjacent sugarcane plots being variable in height resulting in differential shading effects that were not consistent throughout the study.

A replication effect observed for leaf weight was most likely caused by experimental error as leaves of harvested entries had begun to senesce and many had fallen off of the stalk prior to weights being taken (Table 16). This shedding of leaf matter prior to weights being recorded was not uniform as entries were carried varying distances based on their placement in the field which allowed more handling of some entries than others. This assumption is strengthened by the high percentage of variance attributed to error for leaf weight.

For the remaining traits, observed variance due to genotype was over 50% and was over 80% for brix and sucrose measurements (Table 16). While the residual error for these traits was also high it should be noted that due to lack of a replication effect any variance due to replication although not significant was most likely expressed as error. For the sugar related traits brix and sucrose repeatability was estimated at a surprising 72 and 71%, respectively, and repeatability for juice yield was 67%. While variation across multiple environments was not calculated in these estimates it does allow prediction of trait recovery from intergeneric hybrid combination in future studies. However, this estimate does not indicate how interaction of these traits influence plant phenotype as the estimates were made for individual traits across F_1 hybrids sharing only one common parent.

Correlation Analysis

Positive correlations of juice yield with total biomass, stalk and leaf yields and number of stalks were observed (Table 17). Associations of height with internode length, total biomass and stalk yield was also observed. These positive correlations between yield components were expected as was the associations of height with vegetative yield components and internode length. It is logical to assume that taller plants will have longer

internodes and in turn produce more vegetative biomass as was observed in this study. Height however was not associated with leaf yield, indicating that high residual error for leaf yield, likely due to experimental error (Table 16), reduced the ability to accurately determine leaf yield correlations.

In associations of composition traits, ash content was negatively correlated with sucrose but positively correlated with lignin and xylan (Table 18). These results do not coincide with correlations found for photoperiod sensitive RILs (discussed in Chapter 3) as no correlation was found between ash and sucrose in any environment regardless of day length and in the long day environment, a negative association between ash and lignin was observed (Tables 11, 12 and 13). A highly significant association was detected between xylan and lignin in the intergeneric hybrids which is typical between the structural carbohydrates as was observed in the photoperiod sensitive RILs (Tables 11, 12 & 13); as well as by Murray et al., (2008) and Felderhoff et al., (2012). What is unusual within the intergeneric hybrids is that glucan was not associated with any of these compounds. Concurrently, a negative association with sucrose is logical as increases in sugar will, by definition, reduce the relative proportions of other components on a percentage basis. Differences in significant correlations between the intergeneric hybrids discussed herein and previous studies on sorghum RILs is to be expected as RILs shared a similar lineage and the hybrids tested herein are of more diverse genetic backgrounds.

Correlations of phenotypic traits with compositional components indicate a high level of dissociation between the two (Table 19). Brix and bagasse moisture percentage appear to be the most influential on plant composition as most correlations to compositional components were observed for these phenotypic traits (Table 19). Both were positively

correlated with ash, yet inversely associated with sucrose and lignin. These inverse correlations of brix and bagasse moisture percentage are not surprising as brix and bagasse moisture percentage are negatively associated with each other. Brix was negatively associated with xylan content and was the only agronomic trait associated with this compositional component. The positive association of ash and leaf yield cannot be readily explained as leaf matter was not included in this compositional analysis and could not have affected the composition estimates. Glucan was positively associated with stalk yield, height, and internode length and negatively correlated with bagasse moisture percentage. This again was expected as bagasse moisture was negatively associated with height and internode length (Table 17). Overall, the number of significant correlations between phenotypic traits and compositional components was much lower than associations among phenotypic and compositional traits. This may be the result of increased genotypic influence on trait expression due to very diverse genetic backgrounds of the individuals being tested.

In comparison with other studies analyzing phenotypic and composition correlations the results herein are both consistent and inconsistent (Murray et al., 2008; Felderhoff et al., 2012; Stefaniak et al., 2013) as where results between those studies. In addition, the lines being examined in this study are unique in that they are sorghum/sugarcane hybrids which have never been characterized in this way before. It is not necessarily logical to expect them to behave like either parent nor should methods of measurement developed for either parent necessarily be appropriate for the sorcane hybrids (i.e. calibration curves developed specifically for sorghum or sugarcane). This study contained only a small sample of genetic arrangements available between these two species and measurements were recorded within a single environment. Future analysis of hybrids created from identical parents may yield

varying results. Consequently, these initial observations set a baseline from which further testing and evaluation is needed to verify results and identify better methods for evaluation.

Conclusion

The utility of sorghum × sugarcane hybrids was demonstrated as many phenotypic measurements within the hybrids were similar to those in the sugarcane cultivar (Table 15). This indicates that it is possible to express traits deemed desirable by sugarcane producers in intergeneric hybrids at levels that would be acceptable by industry standards. While these existing hybrids lack commercial utility in their current state, they lay the groundwork for continuing efforts toward their overall improvement.

This study incorporated the use of only one sugarcane cultivar in a single environment yet provides justification and promise for further intergeneric hybrid production and testing. The true value of these and subsequent hybrids will have to be assessed in true breeding analysis in which heritability of traits across generations is determined. These factors as well as others which may arise through continued research will ultimately determine the feasibility of such hybrids.

CHAPTER V

CHARACTERIZATION OF NEW SORGHUM GERMPLASM WITH THE IAP ALLELE

Introduction

While the worldwide human population has doubled over the last half century, grain production has also doubled (Charles et al., 2010). Increases in population not only elevate worldwide demand for food, but for energy as well. Given that energy consumption will increase an estimated 57% from the year 2002 to the year 2025 (Office of the Biomass Program 2005; Rooney et al., 2007), there is clearly a need for increasing productivity of all types of crops, including feed, forage, food, fiber and fuel crops (Rooney et al., 2007).

Biomass is expected to contribute to future energy production but to meet the large tonnages required for biofuel production, it will be necessary to develop and produce crops specifically for bioenergy. These dedicated bioenergy crops will be designed to produce high tonnage for fuel or green chemical production and in addition to high yield, they must be tolerant to both biotic and abiotic stresses, and adaptable to a wide range of production environments (Maqbool et al., 2001; Mathews et al., 2000; Dillon et al., 2007). There must also be an array of different bioenergy crop species as this is the only means to produce biomass on a continual basis through the year. Of these species, both sorghum and sugarcane are being developed as dedicated energy crops in the U.S. for which there are active breeding programs (Ahn and Tanksley 1993; Giussani et al., 2001; Paterson 2008; Paterson et al., 2009).

Based on production, sorghum (*S. bicolor*) is the world's fifth most important grain crop (Paterson, 2008). It has been grown in the U.S. traditionally for use as a feed grain and

forage crop; more recently, over 30% of the U.S. grain sorghum crop is converted into ethanol, making it the second most commonly used grain for U.S.-based ethanol production (Paterson, 2008). Sugarcane (*Saccharum* spp. L.) is grown worldwide for sugar production, accounting for 60% of raw sugar consumed worldwide (Grivet and Arruda 2001; Dillon et al., 2007). A significant amount of sugarcane is now used for ethanol production, especially in Brazil (Goldemberg et al., 2004). Despite decades of breeding and research resulting in high yielding, widely adapted varieties, much diversity still exists within both of the species indicating readily available opportunities to further improve plant performance (Wu et al., 2004; Abu Assar et al., 2005; Deu et al., 2006; Kayode et al., 2006; Dillon et al., 2007; Grivet and Arruda 2001).

Diversity within a domesticated crop species has been used in crop improvement programs for over a century. It remains an integral tool to the plant breeding programs of the species. To effectively utilize this genetic diversity, breeders have long relied on hybridization within species (Kuhlman et al., 2008; Hodnett et al., 2010). However, interspecific and even intergeneric hybridization has been used to introduce genetic diversity that is not available in the primary gene pool (Sharma, 1995; Price et al., 2006; Hodnett et al., 2010).

Attempts to introgress traits between *Saccharum* and sorghum began in 1929 when Venkatram attempted to produce early maturing sugarcane (Thomas and Venkatraman 1930). Further efforts were made toward the production of such hybrids in the 1970s when tropical sorghum breeders sought to introgress shoot-fly resistance from sugarcane into sorghum (Young 1972; Gupta et al., 1978). Gupta et al. (1978) reported successful production of *Saccharum* × *Sorghum* hybrids but ultimately trait introgression could not be confirmed

(Gupta et al., 1978). Nair (1999) reported the first confirmed *Sorghum* × *Saccharum* hybrid when five hybrid seedlings were recovered from 3,670 pollinations. Unfortunately, the hybrids produced were limited in both overall number and breeding value.

The limited production of *Sorghum* × *Saccharum* hybrids are the result of reproductive barriers (Hodnett et al., 2005). Removing or overcoming these reproductive barriers is necessary in achieving intergeneric hybridization (Hodnett et al., 2005). The *S. bicolor* mutant allele *iap* (Inhibition of Alien Pollen) reduces some of these reproductive barriers (Hodnett et al., 2005; Price et al., 2006; Kuhlman 2008; Bartek et al., 2012). Introgression of *iap* into BTx623*ms3*, an unreleased, elite, grain sorghum parental line developed by the Texas A&M Experiment station which segregates for genetic male-sterility led to the development of Tx3361*ms3* (Kuhlman and Rooney 2011).

Using Tx3361*ms3*, Hodnett et al. (2010) produced 14,141 seed from 252 Tx3361 × *Saccharum* crosses over a three year period. Seed set was as high as 53% for a single *Saccharum* pollinator and successful hybridization appeared to be male genotype specific (Hodnett et al., 2010). While these F_1 hybrids did not produce viable seed, the ability to produce intergeneric *Sorghum* × *Saccharum* hybrids provides new opportunities for genetic improvement in both species (Hodnett et al., 2010).

The segregation of sterility in Tx3361*ms3* limits its utility in producing large numbers of intergeneric pollinations as it requires constant screening for sterile phenotypes prior to pollination and constant inter-mating to maintain segregation for male sterility. Tx3361*ms3* was sterilized using A1 cytoplasm to create a genetically pure, sterile female which does not segregate for male sterility. The resulting female, designated A.Tx3361, is an unreleased line

developed by the Texas A&M sorghum breeding program. This allowed the production of large quantities of sterile female seed for use in creating large numbers of hybrid crosses.

Intergeneric hybrids created utilizing Tx3361 lack sufficient levels of many traits deemed desirable for sugar production (Chapter 4). Development of an *iap* female genotype containing genes desirable in sugar production will allow the recovery of these traits in hybrid progeny and reduce the loss of desirable traits in resulting hybrids whether intergeneric, interspecific, or intraspecific.

Introgression of *iap* into a more assorted group of elite genotypes will increase the diversity available for creating sweet hybrids and expression of hybrid vigor for sugar producing traits. The objective of this research was to introgress the *iap* allele from Tx3361 into multiple seed parent sweet sorghum genotypes. Following introgression, sterilization of *iap* genotypes will provide stable breeding lines not segregating for sterility and allow large numbers of intergeneric and interspecific crosses without concern of self-pollination.

Materials and Methods

Plant Material

Sorghum Tx3361*ms3* (Kuhlman and Rooney 2011) was used as the source of the *iap* allele and several unreleased sweet sorghum seed parent lines in the Texas A&M Sorghum Breeding Program (B05035, B05038 and B05039) were the source for stalk juiciness and sugar concentrations. A.Tx3361 served as the female parent during sterilization of selected lines being developed. (*Zea mays* L.) 'Kandy Korn' was used as a pollen donor during cytological screening of lines presumed to contain the *iap* allele.

In summer 2008 reciprocal pollinations were made between BTx3361*ms3* and selected sweet sorghum genotypes in College Station, TX. F_1 seed was harvested following grain maturation and planted that fall in Weslaco, TX. Self-pollination was utilized to advance desired genotypes from the F_1 through F_5 generations using pedigree selection breeding methodology. Evaluation and advancement occurred at College Station and Weslaco, TX selecting in each generation for desirable agronomic traits specific to the non *iap* sorghum type (i.e. sweet) used to develop progeny. Lines selected and advanced to F_5 were entered into standard backcross sterilization procedures used by the Texas A&M sorghum breeding program and self-pollinated to generate F_6 lines. An unreleased and cytoplasmic male sterile version Tx3361 (A.Tx3361) was used as the seed parent to initiate sterilization. BC_0F_1 progeny were grown in 2011 at College Station. Throughout sterilization, paired crossing was used to produce A/B pairs. In the fall of 2011 at Weslaco, F_6 pollinators used to create BC_0F_1 progeny were backcrossed to paired BC_0F_1 genotypes to derive BC_1F_1 progeny and selfed to generate F_7 lines.

Because early selection was based on agronomic phenotype rather than presence of the *iap* allele, the population likely contained genotypes that were *Iap* which had to be selected against. To remove the lines that possessed dominant *Iap*, BC_0F_1 progeny were screened for recessive *iap* using flanking markers, developed at Texas A&M University (W.L. Rooney and P.E. Klein), in the fall of 2011. Fresh leaf tissue from BC_0F_1 seedlings was collected and DNA was extracted using the FastDNA SPIN Kit (MP Biomedicals, Santa Ana, CA. Quantification of DNA was performed with a Qubit fluorometer (Life Technologies Corp., Carlsbad, CA) and marker analysis was performed on a Hitachi 3130x1

genetic analyzer (Applied Biosystems, Foster City, CA). Individuals not containing *iap* marker sequences were discarded.

Following marker analysis of recessive *iap,* phenotypic confirmation of each genotype was based on pollen tube observations. Individuals putatively positive for the *iap* allele were pollinated with *Z. mays* var. 'Kandy Korn' following procedures used by Bartek et al. (2012). Pistils were processed according to a modified version of that described by Kho and Baer (1968). Twenty four hours after pollination, pollinated florets were harvested into vials containing 3:1 (95% ethanol: glacial acetic acid) fixative for a minimum of four days. Prior to processing, pistils were extracted from florets and placed overnight into vials containing 0.8M NaOH for a minimum of ten hours, then saturated in a solution containing 0.025% (w/v) aniline blue and 0.1M $K_2 PO_4$ for approximately thirty minutes in the dark. Pistils were then placed on microscope slides in a 1:1 (0.1M K_2PO_4: glycerol) mounting medium and gently covered with a 24 × 50mm cover slip. Observation of specimens was performed using a Zeiss Universal II fluorescent microscope (Carl Zeiss Inc., Gottingen, Germany) equipped with 10X, 25X, and 40X Neofluor objectives, a mercury arc lamp, an excitation 390- to 420-nm bandpass filter, and a 450-nm longpass emission filter. A minimum of twenty-four pistils was observed for each genotype. Genotypes having any pollen tubes growing to the base of the style in at least one of the twenty-four pistils observed were confirmed homozygous for *iap*.

Following *iap* screening in BC_0F_1progeny, BC_1F_1 genotypes containing *iap* were evaluated in College Station in summer 2012 and selected based on agronomic desirability which included acceptable plant height, exertion, maturity, uniformity, brix, minimal lodging and overall agronomic desirability. Selected lines were backcrossed to A.Tx3361 to derive

BC_2F_1 progeny and self-pollinated to create F_8 lines. Seventeen BC_2F_1 genotypes selected at College Station in 2012 were planted August 9, 2012 in Weslaco, TX for evaluation of agronomic traits. Data from the Weslaco nursery was taken on November 6, 2012. Brix readings were recorded in lines developed from sweet sorghum genotypes B05035, B05038, and B05039, as well as Tx3361, to determine relative amount of soluble sugars using a PAL-1 digital pocket refractometer (ATAGO Co., lTD; Itabashi-ku, Japan). In addition to brix values, height, maturity and uniformity were also recorded in all BC_2F_1 lines and Tx3361. Height was recorded in centimeters as the total distance from the base of the plant to the top of the panicle. Maturity was recorded as the number of days after planting in which at least 50% of plants within each plot were at 50% anthesis and uniformity was quantified into a 1 through 9 rating scale in which a value of 1 represented complete uniformity and 9 represented full segregation for phenotypic traits.

Results and Discussion

Of the fourteen genotypes evaluated in Weslaco in the fall of 2012, twelve were derived from the cross of Tx3361*ms3*/B.05035, one was derived from Tx3361*ms3*/B.05038 and one was from the cross of B.Tx3361*ms3*/B.05039 (Table 20). Of the lines tested, only two had higher brix readings than that recorded in Tx3361 (17.1) (Table 20). However, the juicy ratings on the stalks indicate that all of these selections have juicier stalks than Tx3361 and therefore, the relative sugar yields will be higher in them than in Tx3361 which is a dry stalk. The brix readings of 17.2 and 17.6 were recorded in 12WF1967 and 12WF1959, respectively, and both lines were derived from the pedigree of Tx3361*ms3*/B05035.

Flowering time of all genotypes was within three days of Tx3361 which flowered 57 d after planting (Table 20). Genotype 12WF1963 flowered 60 d after planting and was the latest recorded flowering time in the study. Height of genotypes ranged from 91.4cm recorded for 12WF1949 to 167.4cm recorded for 12WF1973 (Table 20). The height of Tx3361 was 121.9cm which indicates a diverse range of heights between the genotypes being developed in comparison to Tx3361. Lodging data was not recorded as it was selected against during development of individual lines and did not differ from Tx3361, in which lodging is not problematic.

Conclusion

While these lines were selected for the *iap* allele for use in wide hybrid production, improvement of brix value and height desirable for sweet sorghum hybrid production allows a more diverse use of developed genotypes. Of the lines tested and relative to Tx3361, 12WF1959 appears to be the best line for use in hybrid production. The increase in brix value over Tx3361 and improved stalk juiciness rating will be valuable in maintaining or increasing sugar yields in sorghum/sugarcane hybrids. These improved traits are applicable to hybrid production whether it is intraspecific within *S. bicolor* or intergeneric with sugarcane.

Further evaluation of these lines is required to establish statistical validity. The testing done to date was based on single line observations (hence not testable statistically). Testing which developed line is best suited for hybrid combination will need to be assessed to confirm this assumption as other traits important to hybrid production were not evaluated in this study such as juice volume and combining ability.

These genotypes tested in hybrid combination with other *S. bicolor* accessions as well as sugarcane will help determine the overall combining ability of the developed lines. An important aspect to be considered when selecting from these lines for use in intergeneric hybrid production is the genotype specific interaction between the *iap* genotypes and selected male parents as noted by Bartek et al. (2012). Until now only a single source of *iap* germplasm has been available to test this interaction. With the development of these lines breeders now have a larger germplasm pool from which to test the genotypic effect of the *iap* allele. Determination as to the effect of the female parent on reducing incompatibility between divergent and alien species may provide valuable insight toward increasing and improving wide hybrid production.

CHAPTER VI

CONCLUSION

Although numerous QTL studies have been completed in sorghum, the study herein is the first report to identify QTL influencing biomass and plant composition in a completely photoperiod sensitive population. Analysis in a day-length sensitive background is essential to eliminate the inherent effect of maturity on both yield and composition. It also implies that testing in variable day lengths is critical to understand other effects relative to the interaction of day length, environment and genotype.

It was found that the varying day lengths and its influence on the ability to detect QTL were not consistent. More phenotypic QTL were detected in the transitional day length environment than in environments which prevented or masked physiological maturity. The inverse was true in identifying QTL influencing composition as many more QTL were detected in the short and long day environments in which maturity effects were eliminated or minimized, respectively. The decreasing day lengths in Weslaco undoubtedly increased phenotypic variation and thus allowed higher detection of QTL in that environment. The reduction in phenotypic variance in the long and short day environments of College Station and Puerto Rico decreased the ability to detect these same QTL within the same population. The decrease in phenotypic variation in the long and short day environments did not however appear to reduce the ability to detect compositional QTL. Twice as many compositional QTL than phenotypic QTL were identified in the long and short day environments compared to the transitional day lengths of Weslaco. This suggests that genetic variance for

composition can be maintained in the absence of maturity effects while expression of phenotypic maturity elevates the genetic expression of phenotypic characteristics.

The findings of day length influence on plant expression of both phenotype and composition loci not only increases our understanding of these two factors but provides future avenues of research for determining the true causes of phenotypic variances observed in photoperiod sensitive sorghum. Further identification of whether QTL expression is due to individual environment or genetic factors, or a combination of the two will provide useful information within sorghum and may be applicable to other crops containing photoperiod sensitive genotypes such as sugarcane.

Comparison and analysis of sorghum × sugarcane hybrids provides opportunities to test the potential of these intergeneric hybrids as a new crop. As the *iap* allele increases the ability to produce intergeneric hybrids between sorghum and sugarcane, characterization of these hybrids improves our ability to make definitive decisions regarding potential uses for this new crop.

The comparison of intergeneric hybrids to elite sugarcane germplasm indicates that these new hybrids do in fact exhibit traits of economic value to commercial sugarcane producers. Furthermore, the ability to recover these desirable traits at high levels establishes its relevance as a novel crop worth further study. The true utility of sorghum × sugarcane hybrids has yet to be determined but the infancy of its study indicate much room for improvement exists in the near future. The ability to produce large quantities of hybrids facilitated by the *iap* allele will allow broader testing of existing sugarcane germplasm in combination with sorghum. The high level of repeatability exhibited by these first generation

progeny indicate that replication of high valued traits is theoretically attainable in future hybrids.

Evaluation of genetic stability of new generations of sorghum × sugarcane hybrids will allow breeders to determine its feasibility within sugarcane and sorghum breeding programs alike. Although it was not reported within this research, variation within propagated intergeneric genotypes was observed under field conditions indicating instability of hybrid genotypes. Until the stability of these genotypes is established their true value toward trait introgression between the species cannot be determined.

It is essential that new generations of intergeneric hybrids be evaluated in comparison to their parental counterparts to ensure that new hybrids will improve or surpass what is currently available to researchers and producers. While it is understood that intergeneric hybrids will most likely not replace sorghum or sugarcane, the hybrids themselves may provide an end use function and at a minimum, serve as new sources of diversity within breeding programs for both species. Determination of this will depend on the ability of researchers to produce and characterize even larger numbers of intergeneric hybrids using more diverse sources of parental germplasm than was presented herein.

Development of more diverse sorghum genotypes containing the *iap* allele allows selection of parental lines capable of producing intergeneric hybrids containing traits of interest, depending on their potential end use, and provides larger sources of hybrid combinations to select from. The sweet sorghum genotypes containing *iap* may be used to create sorghum × sugarcane hybrids yet maintain percentages of sugar similar to sugarcane, surpassing those of intergeneric hybrids facilitated by Tx3361. Sweet sorghum genotypes

containing *iap* provide a more diverse genetic background for intergeneric crosses than what was previously available when Tx3361 was the only available *iap* sorghum genotype. This will provide a broader range of desirable agronomic traits to be incorporated into the sugarcane genome which may behave differently than those of Tx3361 in hybrid combination with sugarcane.

Continued efforts to identify genetic and environmental factors affecting plant phenotypic expression will aid in the ability to develop, evaluate, and produce agronomically superior genotypes. Corollary analysis of improved genotypes, as parental lines and in hybrid combination, will provide fundamental insight for the next step to be taken in the areas of trait discovery, implementation, and wide hybridization.

The results from studies provided herein were the first of their kind to be conducted. Yet different in nature they all aim to broaden our understanding of the sorghum genome and its uses as a cultivated crop. All studies conducted in this dissertation were facilitated through previous research which posed broad scientific questions. While some of the questions were answered through the completion of this research it has spawned new ideas for future studies.

REFERENCES

Abu Assar, A.H., R. Uptmoor, A.A. Abdelmula, M. Salih, F. Ordon, W. Friedt. 2005. Genetic variation in sorghum germplasm from Sudan, ICRISAT, and USA assessed by simple sequence repeats (SSRs). Crop Sci. 45:1636-1644.

Ahn, S. and S.D. Tanksley. 1993. Comparative linkage maps of the rice and maize genomes. Proc. Natl. Acad. Sci. 90:7980-7984.

Bartek, M.S., G.L. Hodnett, B.L. Burson, D.M. Stelly, W.L. Rooney. 2012. Pollen tube growth after intergeneric pollinations of *iap*-homozygous *Sorghum bicolor.* Crop Sci. 52:1553-1560.

Berding, N., M. Hogarth, M. Cox. 2004. Plant improvement of sugarcane. P.20-49. *In*. G.L. James (ed.) Sugarcane. Blackwell Science Ltd. Ames, IA.

Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci. 48:1649-1664.

Bowers, J.E., C. Abbey, S. Anderson, C. Chang, X. Draye, A.H. Hoppe, R. Jessup, C. Lemke, J. Lennington, Z. Li, S. Liu, L. Luo, B. Marler, R. Ming, D. Qiang, K. Reischman, S. Shulz, D. Skinner, Y. Wang, S. Kresovich, K. Schertz, A. Paterson. 2003. A high-density genetic recombination map of sequence tagged sites for sorghum, as a framework for comparative, structural and evolutionary genomics of tropical grains and grasses. Genetics. 165:367-386.

Charles, H., J. Godfray, J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, J.F. Muir, J. Pretty, S. Robinson, S.M. Thomas, C. Toulmin. 2010. Food Security: The challenge of feeding 9 billion people. Science 327:812-817.

Childs, K.L., F.R. Miller, M.M. Cordonnier-Pratt, L.H. Pratt, P.W. Morgan, J.E. Mullet. 1997. The sorghum photoperiod sensitive gene, Ma_3, encodes a Phytochrome B^1. Plant Phys. 113:611-619.

Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics. 138:963-971.

Corn, R.J. 2009. Heterosis and composition of sweet sorghum. PhD. Diss., Texas A&M Univ., College Station.

Dahlberg, J., E. Wolfrum, B. Bean, and W.L. Rooney. 2011. Compositional and agronomic evaluation of sorghum biomass as a potential feedstock for renewable fuels. J. Biobased Mater. Bioenergy 5:507-513.

Deu, M., F. Rattunde, J. Chantereau. 2006. A global view of genetic diversity in cultivated sorghums using a core collection. Genome. 49:168-180.

Dillon S.L., F.M. Shapter, R.J. Henry, G. Cordeiro, L. Izquierdo, L.S. Lee. 2007. Domestication to crop improvement: genetic resources for *Sorghum* and *Saccharum* (Andropogoneae). Ann. Bot-London. 100:975-989.

Felderhoff, T.J., S.C. Murray, P.E. Klein, A. Sharma, M.T. Hamblin, S. Kresovich, W. Vermerris, W.L. Rooney. 2012. QTLs for energy- related traits in a sweet × grain sorghum [*Sorghum bicolor* (L.) Moench] mapping population. Crop Sci. 52:2040-2049.

Feltus, F., H. Hart, K. Scherta, A. Casa, S. Kresovich, S. Abraham, P.E. Klein, A. Paterson. 2006. Alignment of genetic maps and QTLs between inter- and intra- specific sorghum populations. Theor. Appl. Genet. 112:1295-1305.

Giussani, L.M., J.H. Cota- Sanchez, F.O. Zuloaga, E.A. Kellogg. 2001. A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C_4 photosynthesis. Amer. J. Bot. 88:1993-2012.

Goldemberg, J., S.T. Coelho, P.M. Nastari, O. Lucon. 2004. Ethanol learning curve- the Brazilian experience. Biomass and Bioenergy 26:301-304.

Grivet, L. and P. Arruda. 2001. Sugarcane genomics: depicting the complex genome of an important tropical crop. Curr. Opin. Plant Biol. 5:122-127.

Gupta, S.C., J.M.J. DeWet, J.R. Harlan. 1978. Morphology of *Saccharum-Sorghum* hybrid derivatives. Amer. J. Bot. 65:936-942.

Haley, C.S., and S.A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315-324.

Hodnett, G.L., B.L. Burson, W.L. Rooney, S.L. Dillon, and H.J. Price. 2005. Pollen-pistil interactions result in reproductive isolation between *Sorghum bicolor* and divergent *Sorghum* species. Crop Sci. 45:1403-1409.

Hodnett, G.L., A.L. Hale, D.J. Packer, D.M. Stelly, J. da Silva, and W.L. Rooney. 2010. Elimination of a reproductive barrier facilitates intergeneric hybridization of *Sorghum bicolor* and *Saccharum*. Crop Sci. 50:1188-1195.

Kayode, A.P.P., A.R. Linnemann, M.J.R. Nout, J.D. Hounhouigan, T.J. Stomph, M.J.M. Smulders. 2006. Diversity and food quality properties of farmers' varieties of sorghum from Benin. J. Sci. Food Agric. 86:1032-1039.

Kuhlman, L.C., B.L. Burson, P.E. Klein, R.R. Klein, D.M. Stelly, H.J. Price, W.L. Rooney. 2008. Genetic recombination in *Sorghum bicolor* × *S. macrospermum* interspecific hybrids. Genome 51:749-756.

Kuhlman, L.C. and W.L. Rooney. 2011. Registration of Tx3361 sorghum germplasm. J. Plant Reg. 5:133-134.

Laurie, D.A. and M.D. Bennett. 1989. Genetic variation in sorghum for the inhibition of maize pollen tube growth. Ann. Bot. (Lond.) 64:675-681.

Lin, Y.R., K.F. Schertz, A.H. Paterson. 1995. Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. Genetics. 141:391-411.

Maqbool, S.B., D. Prathibha, M.B. Sticklen. 2001. Biotechnology: Genetic improvement of Sorghum (*Sorghum bicolor* (L.) Moench). In Vitro Cell. Dev. Biol.-Plant 37:504-515.

Mathews, S., R.C. Tsai, E.A. Kellogg. 2000. Phylogenetic structure in the grass family (Poaceae): Evidence from the nuclear gene phytochrome B. Amer. J. Bot. 87:96-107.

Menz, M.A., R.R. Klein, J.E. Mullet, J.A. Obert, N.C. Unruh, P.E. Klein. 2002. A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP®, RFLP and SSR markers. Plant Mol. Biol. 48:483-499.

Monti, A., N.D. Virgilio, G. Venturi. 2008. Mineral composition and ash content of six major energy crops. Bio. And Ener. 32:216-223.

Murphy, R.L., R.R. Klein, D.T. Morishige, J.A. Brady, W.L. Rooney, F.R. Miller, D.V. Dugas, P.E. Klein, J.E. Mullet. 2011. Coincident light and clock regulation of pdeudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. PNAS. 108:16469-16474.

Murray, S.C., A. Sharma, W.L. Rooney, P.E. Klein, J.E. Mullet, S.E. Mitchell, S. Kresovich. 2008. Genetic improvement of Sorghum as a Biofuel Feedstock: I QTL for Stem Sugar and Grain Nonstructural Carbohydrates. Crop Sci. 48:2165-2179.

Nair, N.V. 1999. Production and cyto-morphological analysis of intergeneric hybrids of *Sorghum* × *Saccharum*. Euphytica 108:187-191.

Paterson A.H. 2008. Genomics of *Sorghum*. Int. J. Plant Genomics. doi:10.1155/2008/362451.

Paterson, A.H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Poliakov, J. Schmutz, M. Spannagl, H. Tang, X. Wang, T. Wicker, A.K. Bharti, J. Chapman, F.A. Feltus, U. Gowik, I.V. Grigoriev, E. Lyons, C.A. Maher, M. Martis, A. Narechania, R.P. Otillar, B.W. Penning, A.A. Salamov, Y. Wang, L. Zhang, N.C. Carpita, M. Freeling, A.R. Gingle, C.T. Hash, B. Keller, P. Klein, S. Kresovich, M.C. Mccann, R. Ming, D.G. Peterson, R. Mehboob Ur, D. Ware, P. Westhoff, K.F.X. Mayer, J. Messing, and D.S. Rokhsar. 2009. The Sorghum bicolor genome and the diversification of grasses. Nature 457:551-556.

Perez, J., J. Munoz-Dorado. 2002. Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. Int. Microbiol. 5:53-63.

Perlack, R.D., L.L. Wright, A.F. Turhollow, R.L. Graham, B.J. Stokes, D.C. Erbach. 2005. Biomass as a feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply. USDA/DOE, DOE/GO-102005-2135.

Price, H.J., G.L. Hodnett, B.L. Burson, S.L. dillon, D.M. Stelly, W.L. Rooney. 2006. Genotype dependent interspecific hybridization of *Sorghum bicolor*. Crop Sci. 46:2617-2622.

Reddy, B.B.S., S. Ramesh, P.S. Reddy. 2006. Sorghum genetic resources, cytogenetics and improvement. P.309-363. *In*. R.J. Singh and P.P. Jauhar (ed.) Genetic resources, chromosome engineering, and crop improvement. CRC Taylor and Francis, Boca Raton, FL.

Ritter, K.B., D.R. Jordan, S.C. Chapman, I.D. Godwin, E.S. Mace, C.L. McIntyre. 2008. Identification of QTL of sugar- related traits in a sweet × grain sorghum (*Sorghum bicolor* L. Moench) recombinant inbred population. Mol. Breeding. 22:367-384.

Rooney, W.L., 2000. *Genetic and Cytogenetics*. p.261-308. *In*. C.W. Smith and R.A. Frederiksen(ed.) *Sorghum* origin, history, technology, and production. John Wiley & Sons Inc., New York, NY.

Rooney, W.L. 2004. Sorghum Improvement- Integrating traditional and new technology to produce improved genotypes. Adv. In Agr. 83:37-109.

Rooney, W.L., J. Blumenthal, B. Bean, J.E. Mullet. 2007. Designing sorghum as a dedicated bioenergy feedstock. Biofuels, Bioproducts and Biorefining. 1:147-157.

Rosenow, D.T. and J.A. Dahlberg. 2000. Collection, conversion, and utilization of sorghum. P.309-328. *In* C.W. Smith (ed.) Sorghum: Origin, history, technology, and production. John Wiley & Sons, New York, NY.

Office of the Biomass Program. 2005. Multi year program plan 2007-2012. United States Department of Energy. http://www1.eere.energy.gov/biomass/pdfs/mypp.pdf.

Sharma, H.C. 1995. How wide can a wide cross be? Euphytica. 82:43-64.

Stefaniak, T.R., J.A. Dahlberg, B.W. Bean, N. Dighe, E.J. Wolfrum, W.L. Rooney. 2012. Variation in biomass composition components among forage, biomass, sorghum-Sudangrass, and sweet sorghum types. Crop Sci. 52:1949-1954.

Stephens, J.C., F.R. Miller, D.T. Rosenow. 1967. Conversion of alien sorghums to early combine genotypes. Crop Sci. 7:396.

Theander, O. and E. Westerlund. 1993. Quantitative analysis of cell wall components. In: Forage, Cell Wall Structure and Digestibility. ASA-CSSA-SSSA. Madison, WI. Pg 83-104.

Thomas, R. and T.S. Venkatraman. 1930. Sugarcane-sorghum hybrids. Agric. J. India. 25:164.

Wolfrum, E., C. Payne, T. Stefaniak, W. Rooney, N. Dighe, B. Bean, J. Dahlberg. 2013. Multivariate Calibration Models for Sorghum Composition using Near- Infrared Spectroscopy. US Dept. Ener.

Wu, J.L., C.J. Wu, C.L. Lei, M. Baraoidan, M.R.S. Madamba. 2004. Sorghum diversity evaluated by simple sequence repeat (SSR) markers and phenotypic performance. Plant Prod. Sci. 7:301-308.

Young, W.R. 1972. Sources of resistance to sorghum shoot-fly (*Atherigona varia soccata* Rond.). P. 168-179. *In* M.G. Jotwani and W.R. Young (eds.) Control ofsorghum shoot-fly. Oxford and I.B.H. Pub. Co., New Delhi.

Zhanyou, X., D. Zhang, J. Hu, X. Zhou, X. Ye, K.L. Reichel, N.R. Stewart, R.D. Syrenne, X. Yang, P. Gao, W. Shi, C. Doeppke, R.W. Sykes, J.N. Burris, J.J. Bozell, Z.M. Cheng, D.G. Hayes, N. Labbe, M. Davis, C.N. Stewart, J.S. Yuan. 2009. Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. BMC Bioinformatics (Suppl 11):S3. doi:10.1186/1471-2105-10-S11-S3.

APPENDIX

Table 1. Phenotypic measurements of parental lines R07020, R07018 and F_5 population for all three environments (College Station, Weslaco, and Puerto Rico).

Trait	2011 College Station					2011 Weslaco Fall					2012 Puerto Rico				
	R07020	R07018	F_5 mean†	F_5 range	LSD¶	R07020	R07018	F_5 mean	F5 range	LSD¶	R07020	R07018	F_5 mean†	F_5 range	LSD¶
Height, cm	427	414	451.99(45)	339-530	82.1	342	355	334(14)	298-373	22.2	237	246	251(15)	215-280	19.6
Exertion, cm	na‡	na‡	na‡	na‡	na‡	0	0	0.6(1.3)	0-6	1.7	6.5	8	7(1.8)	3-12.5	2.9
Days to Flower, d	na‡	na‡	na‡	na‡	na‡	95	90	88(4)	80-97	2.9	61	64	61(2)	56-65	2.1
Harvest Yield, Mg ha^{-1}	73.2	65.9	61.9(20.3)	8.6-105.8	40.2	57.5	64	65.9(11.8)	44.5-97.6	8.6	58.5	47.8	55.2(10.7)	28.3-78.3	21.2
Stalk Yield, Mg ha^{-1}	57.4	46.9	51.5(17.2)	6.6-89.5	34.1	50.6	52	56.8(8.7)	40.5-79.1	8.5	na‡	na‡	na‡	na‡	na‡
Leaf Yield, Mg ha^{-1}	15.4	18.8	10.5(3.7)	1.9-18.2	8.5	6.9	11.7	8.6(2.8)	3.7-16.6	5.4	na‡	na‡	na‡	na‡	na‡
Panicle Yield, Mg ha^{-1}	na‡	na‡	na‡	na‡	na‡	2.9	3.2	2.6(1.5)	0.1-5.5	1.7	7.6	2.1	8.3(3.6)	1.7-16.4	5.3
Biomass Moisture, %	69	70	72(5)	57-85	9.6	61	60	62(4)	52-72	6.6	65	68	69(3)	62-76	5.2
Juice Yield, L ha^{-1}	4303	5667	12298(5252)	734-25282	8722	5052	4921	4561(1554)	1509-8464	3013	na‡	na‡	na‡	na‡	na‡
Brix, %	7.7	8.8	7.8(1)	5.5-10.2	2.3	7.7	8.7	8.6(1.6)	3.9-12.8	2.3	7.6	9.7	8.5(1.1)	5.6-11.6	1.9
Internode Length, cm	21	16	19(3)	12-27	5.2	26	24	24(3)	17-29	3.5	27	26	27(2)	22-31	3.2
Stem Diameter, mm	16	17	18(2)	11-23	3.9	18	13	15(2)	10-21	4.5	10	12	11(1)	8-15	3.1

† Standard deviation reported in parenthesis.

‡ Data not available or trait not expressed.

§ Bagasse Moisture % = 100 × (1-(pressed stalked dry weight/ pressed stalk wet weight))

¶ Least significant difference calculated for each trait at α=0.05

Table 2. Relative percentage of calculated variances attributed to genetic (Gen), environmental (Env), genotype × environment interactions (Gen × Env) and other experimental factors.

Trait	College Station		Weslaco		Puerto Rico		All Locations			
	Gen.	Residual Error	Gen.	Residual Error	Gen.	Residual Error	Gen	Env	Gen × Env	Residual Error
	%		%		%		%			
Height	56.41***	43.58	68.71***	31.29	81.45***	18.55	0.88	94.16***	1.97***	2.99
Exertion	na†	na†	66.15***	33.85	78.59***	21.41	0.39**	98.44**	0.63***	0.55
Days to Flower	na†	na†	93.44***	6.56	86.99***	13.01	17.83	46.26***	35.77***	0.15
Harvest Yield	78.06***	21.94	57.53***	42.47	65.92***	34.08	0.67	96.37***	1.84***	1.12
Stalk Weight	77.27***	22.74	55.65***	44.35	na†	na†	1.01	96.06**	1.80***	1.14
Leaf Weight	78.17***	21.83	63.72***	36.28	na†	na†	0.06	99.79**	0.10***	0.06
Panicle Weight	na†	na†	75.76***	24.24	55.11***	44.88	0.01	99.97**	0.01***	0.02
Biomass Moisture %	62.86***	37.15	62.62***	37.38	75.11***	24.89	1.82*	90.17**	3.58***	4.42
Juice Yield	71.69***	28.31	40.99*	59.01	na†	na†	7.18	70.08***	13.47***	9.28
Brix	64.23***	35.77	65.00***	35.00	52.88***	47.12	0.18**	99.04*	0.30***	0.48
Internode Length	68.74***	31.26	77.47***	22.53	63.08***	36.92	0.78*	96.32**	1.47***	1.43
Stem Diameter	57.20***	42.80	76.43***	23.57	52.44***	47.56	10.29	19.12*	30.87***	39.72

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level.
†Data not available or trait not expressed.

Table 3. Broad sense heritability (H^2) and standard errors for phenotypic trait means measured in College Station, Weslaco, and Puerto Rico as well as combined analysis across all environments.

Trait	College Station	Weslaco	Puerto Rico	All Locations
Height	0.39 ± 0.08	0.52 ± 0.07	0.69 ± 0.05	0.33 ± 0.16
Exertion	na†	0.49 ± 0.07	0.65 ± 0.05	na†
Days to Flower	na†	0.88 ± 0.02	0.65 ± 0.04	na†
Biomass Yield	0.64 ± 0.05	0.40 ± 0.08	0.49 ± 0.07	0.13 ± 0.05
Stalk Yield	0.63 ± 0.06	0.39 ± 0.09	na†	na†
Leaf Yield	0.64 ± 0.05	0.47 ± 0.08	na†	na†
Panicle Yield	na†	0.61 ± 0.06	0.38 ± 0.08	na†
Biomass Moisture %	0.47 ± 0.07	0.46 ± 0.07	0.60 ± 0.06	0.18 ± 0.12
Juice Yield	0.56 ± 0.07	0.26 ± 0.10	na†	na†
Brix	0.47 ± 0.07	0.48 ± 0.07	0.36 ± 0.08	0.17 ± 0.08
Internode Length	0.52 ± 0.07	0.63 ± 0.06	0.46 ± 0.07	0.69 ± 0.39
Stem Diameter	0.40 ± 0.08	0.62 ± 0.06	0.36 ± 0.08	0.23 ± 0.04

†Data is not available or non-significant source of variation.
‡Combined analysis performed for traits present in all environments.

Table 4. Pearson coefficients of correlation for phenotypic traits measured in College Station.

Trait	Height (cm)	Volume (L/ha-1)	Biomass Yield (Mg/ha-1)	Stalk Yield (Mg/ha-1)	Leaf Yield (Mg/ha-1)	Internode Length (cm)	Stem Diameter (mm)	Plot Moisture (%)
Brix (%)	0.28**	0.25*	0.21*	0.19	0.27**	0.08	0	-0.03
Height (cm)	—	0.22*	0.15	0.16	0.07	0.06	-0.06	-0.07
Juice Yield (L/ha-1)		—	0.46***	0.44***	0.43***	0.1	0.44***	-0.01
Biomass Yield (Mg/ha-1)			—	0.99***	0.85***	0.08	0.26**	0
Stalk Yield (Mg/ha-1)				—	0.78***	0.08	0.24*	0.01
Leaf Yield (Mg/ha-1)					—	0.06	0.26**	-0.03
Internode Length(cm)						—	-0.05	0.04
Stem Diameter (mm)							—	-0.03
Biomass Moisture (%)								—

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level.

Table 5. Correlation of phenotypic traits recorded in Weslaco.

Trait	Height (cm)	Exertion (cm)	Volume (L/ha-1)	Biomass Yield (Mg/ha-1)	Stalk Yield (Mg/ha-1)	Leaf Yield (Mg/ha-1)	Panicle Yield (Mg/ha-1)	Internode Length (cm)	Stem Diameter (mm)	Plot Moisture (%)	Flowering Time (d)
Brix (%)	-0.17	-0.12	-0.13	-0.1	-0.13	0.05	-0.14	0.15	-0.19	-0.06	0.07
Height (cm)	—	0.09	0.03	0	0.05	-0.16	0.04	-0.06	0.21*	0.15	0
Exertion (cm)		—	-0.12	-0.15	-0.11	-0.21*	-0.22*	-0.24*	-0.03	-0.06	-0.41***
Juice Yield (L/ha-1)			—	0.34***	0.38***	0.16	0.18	0.06	0.06	0.02	-0.07
Biomass Yield (Mg/ha-1)				—	0.98***	0.4***	0.6***	0.08	-0.02	-0.1	0.1
Stalk Yield (Mg/ha-1)					—	0.33***	0.54***	0.06	-0.01	-0.07	0.07
Leaf Yield (Mg/ha-1)						—	0.15	0.22*	0.02	0.07	0.26*
Panicle Yield (Mg/ha-1)							—	-0.05	0	-0.12	0.08
Internode Length (cm)								—	-0.03	-0.09	0.2
Stem Diameter (mm)									—	0.18	0.15
Biomass Moisture (%)										—	0.14
Flowering Time (d)											—

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level.

Table 6. Pearson correlations of phenotypic measurements recorded in Puerto Rico.

Trait	Height (cm)	Exertion (cm)	Biomass Yield (Mg/ha-1)	Panicle Yield (Mg/ha-1)	Internode Length (cm)	Stem Diameter (mm)	Plot Moisture (%)	Flowering Time (d)
Brix (%)	0	0.02	0.13	0.029	0.03	0.2*	0.01	-0.11
Height (cm)	—	0.43***	0.2*	0.06	0.25*	0.11	-0.31**	0.05
Exertion (cm)		—	0.05	-0.12	0.19	0.1	-0.13	-0.17
Biomass Yield (Mg/ha-1)			—	0.5***	0.07	0.04	0	-0.07
Panicle Yield (Mg/ha-1)				—	-0.24**	-0.03	-0.04	0.04
Internode Length (cm)					—	-0.12	-0.24*	-0.12
Stem Diameter (mm)						—	0.06	0.06
Biomass Moisture (%)							—	-0.23*
Flowering Time (d)								—

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level.
†Data not available or trait not expressed.

Figure 1. Genetic map developed from the R07018 × R07020 population with entries treated as RILs. Markers coincide with their respective chromosome with chromosomes 4 and 5 representing two linkage groups as reported by JoinMapV4.0. Marker order on chromosomes was determined by the BTx623 sorghum genome assembly sequence (Paterson et al., 2009).

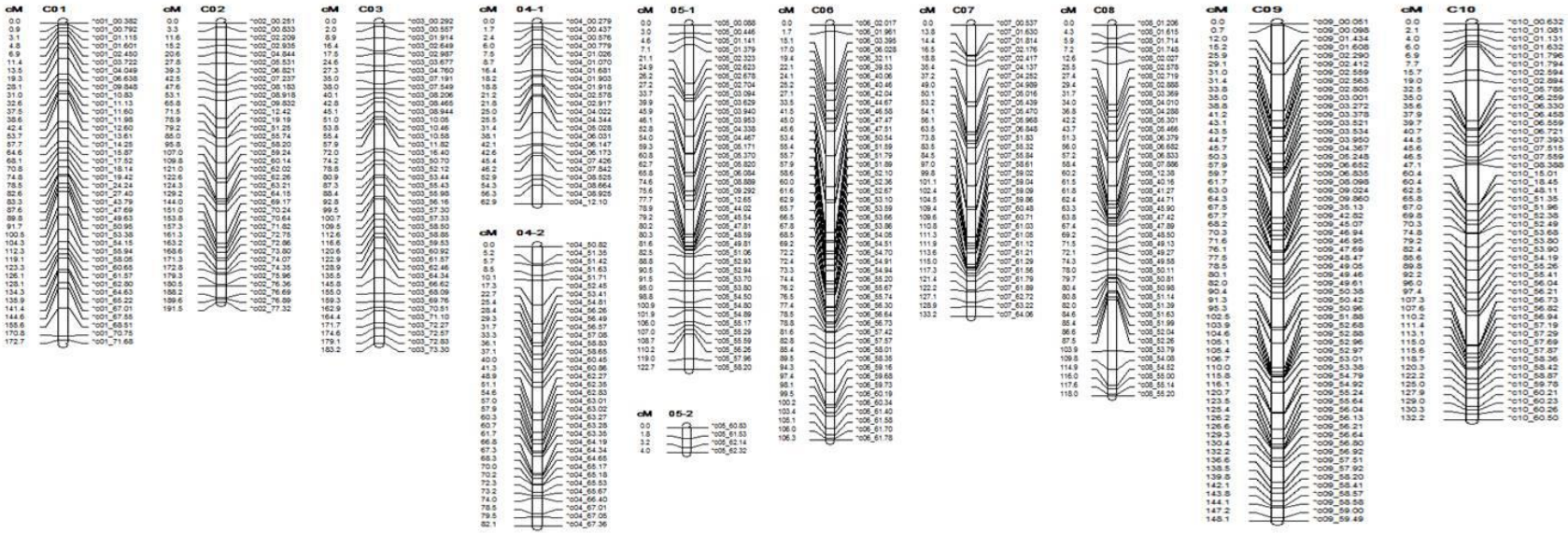

Table 7. Genetic map position and relative effect of phenotypic QTL from collected data recorded in College Station, Weslaco and Puerto Rico. Quantitative traits are separated by environment and chromosome in which they were identified. Genetic location of each trait was determined by the Mega base pair position of identifying markers (Mbp) and peak centimorgan distance (cM) on each chromosome. Contributing parent and effect indicate which parental line contributed alleles that increased trait values and effect is the phenotypic contribution of those alleles. The 2- likelihood of odds interval indicates the centimorgan distance for which significant QTL were detected and R^2 represent the variance explained by each QTL.

Trait	Environment	Chromosome	QTL Location (Mbp)	QTL Peak (cM)	LOD 2 QTL Interval (cM)	Contributing Parent	Effect	R^2	LOD
Juice Yield (mL)	College Station	2	9.82	64.2	54.2-70	R07018	241	0.11	3.3
Internode Length (cm)	College Station	9	6.65	58.9	53.7-59.7	R07020	0.93	0.14	3.5
Stem Diameter (mm)	College Station	3	56.15	92.4	89.4-99.5	R07020	0.75	0.15	4.2
Internode Length (cm)	Puerto Rico	7	61.03	110.8	110.7112.9	R07018	0.75	0.14	3.9
Days to Flower (d)	Puerto Rico	3	56.15	90.4	88.392.8	R07018	0.68	0.1	3.3
Brix (%)	Weslaco	6	44.67	35.5	33.241.5	R07020	0.55	0.11	3.3
Height (cm)	Weslaco	10	56.73	107.3	97.4109.9	R07020	2.64	0.16	4.7
Stalk Yield (kg)	Weslaco	3	2.98	17.4	12.921.1	R07020	0.27	0.09	2.9
Leaf Yield (kg)	Weslaco	10	56.73	107.3	97.3112.5	R07018	0.07	0.14	4.1
Panicle Yield (kg)	Weslaco	2	12.42	71.6	70.174.6	R07018	0.03	0.13	3.8
Panicle Yield (kg)	Weslaco	2	15.25	78.4	74.679	R07018	0.03	0.13	4
Days to Flower (d)	Weslaco	2	73.04	164.5	164.3168.2	R07020	1.79	0.19	6.6

†Phenotypic contribution, noted as effect, is reflective of quantitative increase recorded in a 2m sample.

Table 8. Measurements of compositional traits for parental lines R07020, R07018 and F_5 population for College Station, Weslaco and Puerto Rico.

Trait	College Station					Weslaco					Puerto Rico				
	R07020	R07018	F_5 Mean†	F_5 Range	LSD¶	R07020	R07018	F_5 Mean	F_5 Range	LSD¶	R07020	R07018	F_5 Mean	F_5 Range	LSD¶
Ash, %	7.6	6.2	7(0.6)	5.4-8.4	0.8	7.3	7.4	6.1(0.6)	4.7-7.3	1.2	7.3	7.6	7.8(0.5)	6.4-9.2	0.8
Protein, %	3.4	4	4.2(0.3)	3.3-4.9	0.6	3.27	4.3	3.5(0.4)	2.5-4.5	0.9	2.7	3.4	3.5(0.3)	2.8-4	0.6
Sucrose, %	8.7	9.9	9.6(1.7)	5.7-14.2	3.1	11.5	8.8	11.2(1.8)	7.5-16.3	2.7	4	1	2.3(1.2)	0.6-6.9	1.3
Lignin, %	13.2	12.6	12.5(0.7)	10.6-14.7	1.5	14.7	14.3	14.3(0.8)	12.3-16.2	1.5	15.1	15.6	15.3(0.7)	13.4-16.8	0.9
Xylan, %	16.1	14.9	15(0.6)	13.7-16.8	1.1	16.6	16.5	16.6(0.6)	15.1-17.8	1.0	17.5	17.9	17.6(0.5)	16.2-18.7	0.6
Glucan, %	30.8	30.5	30.9(0.9)	28.1-33.4	1.8	31.2	29.9	30.6(0.9)	28.4-32.8	1.6	33.9	35.1	33.7(0.8)	31.9-36.1	1.3

†Standard deviation reported in parenthesis.

‡ Data not available or trait not expressed.

¶ Least significant difference calculated for each trait at α=0.05.

Table 9. Proportion of calculated variance for compositional traits due to genotype (Gen.), environment (Env.), genotype × environment interaction (Gen. × Env.) as well as other experimental factors

Trait	College Station		Weslaco		Puerto Rico		All Locations			
	Gen.	Residual Error	Gen.	Residual Error	Gen.	Residual Error	Gen	Env	Gen × Env	Residual Error
Ash	80.59***	19.41	62.9***	37.10	76.15***	23.85	0.09***	99.68**	0.09***	0.13
Protein	69.8***	30.20	67.3***	32.67	71.08***	28.92	0.03	99.84*	0.06***	0.07
Sucrose	62.82***	37.18	76.6***	23.40	79.61***	20.39	0.37**	98.38***	0.59***	0.66
Lignin	64.76***	35.24	64.25***	35.75	75***	25	0.09*	99.49**	0.18***	0.24
Glucan	61.75***	38.26	67.79***	32.21	75.15***	24.85	0.16***	99.26**	0.23***	0.35
Xylan	66.64***	33.36	72.46***	27.54	79.38***	20.62	0.06*	99.69**	0.11***	0.13

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level.
†Data not available or trait not expressed.

Table 10. Heritability estimates (H^2) and standard errors for compositional traits measured in College Station, Weslaco, and Puerto Rico and combined analysis of all environments.

Trait	College Station	Weslaco	Puerto Rico	All Locations
Ash	0.68 ± 0.05	0.46 ± 0.07	0.61 ± 0.06	0.19 ± 0.04
Protein	0.54 ± 0.07	0.51 ± 0.07	0.55 ± 0.07	0.02 ± 0.01
Sucrose	0.46 ± 0.07	0.62 ± 0.06	0.66 ± 0.05	0.44 ± 0.12
Lignin	0.48 ± 0.07	0.47 ± 0.08	0.6 ± 0.06	0.07 ± 0.04
Glucan	0.45 ± 0.08	0.51 ± 0.07	0.6 ± 0.06	0.18 ± 0.07
Xylan	0.5 ± 0.07	0.57 ± 0.07	0.66 ± 0.05	0.052 ± 0.03

†Data is not available or trait was not expressed.
‡Combined analysis performed for traits present in all environments.

Table 11. Correlation of compositional traits measured in College Station.

Trait	Ash (%)	Protein (%)	Sucrose (%)	Lignin (%)	Xylan (%)	Glucan (%)
Ash (%)	-	0.37***	-0.19	-0.24*	0.14	-0.2*
Protein (%)		-	-0.3**	-0.04	-0.13	-0.21*
Sucrose (%)			-	-0.64***	-0.74***	-0.24*
Lignin (%)				-	0.84***	0.48***
Xylan (%)					-	0.42***
Glucan (%)						-

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level

Table 12. Pearson correlations of compositional traits measured in Weslaco.

Trait	Ash (%)	Protein (%)	Sucrose (%)	Lignin (%)	Xylan (%)	Glucan (%)
Ash (%)	-	-0.08	0.15	-0.16	-0.14	-0.09
Protein (%)		-	-0.41***	-0.08	-0.07	-0.44***
Sucrose (%)			-	-0.8***	-0.76***	-0.26**
Lignin (%)				-	0.89***	0.55***
Xylan (%)					-	0.46***
Glucan (%)						-

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level.

Table 13. Trait correlations of compositional components measured in Puerto Rico.

Trait	Ash (%)	Protein (%)	Sucrose (%)	Lignin (%)	Xylan (%)	Glucan (%)
Ash (%)	-	0.29**	-0.17	-0.15	0.18	-0.09
Protein (%)		-	-0.19*	0.02	-0.19	-0.27**
Sucrose (%)			-	-0.79***	-0.76***	-0.69***
Lignin (%)				-	0.82***	0.58***
Xylan (%)					-	0.58***
Glucan (%)						-

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level.

Table 14. Compositional QTL identified in College Station, Weslaco and Puerto Rico. QTL genetic location is reported by mega base pair position of identifying markers (Mbp) and peak centimorgan location (cM) on each chromosome. Genetic distances for each QTL are reported as LOD 2 QTL interval. Contributing parent and effect indicate source of alleles and their contribution to quantitative measurement. Variance due to each loci is reported as $R^{2.}$

Trait	Location	Chromosome	QTL Location (Mbp)	QTL Peak (cM)	LOD 2 QTL Interval (cM)	Contributing Parent	Effect	R^2	LOD
Protein	College Station	1	10.83	31	29.1-34.5	R07020	0.12	0.11	3.7
Sucrose	College Station	3	4.76	27.4	24.1-34.4	R07018	0.56	0.09	3.7
Ash	College Station	4	8.29	56.4	53.5-61.4	R07020	0.29	0.08	3.3
Lignin	College Station	4	56.57	31.7	31.3-33.3	R07020	0.28	0.09	3.5
Protein	College Station	4	58.65	37.1	35.3-39.4	R07020	0.22	0.17	5.7
Ash	College Station	5	3.94	45.9	41.1-45.9	R07018	0.24	0.1	3.6
Ash	College Station	5	12.65	77.7	76.4-79.2	R07020	0.24	0.14	4.8
Ash	College Station	6	58.01	85.4	79.1-93.8	R07018	0.2	0.11	4.2
Sucrose	College Station	6	46.58	41.5	34.4-41.5	R07020	0.54	0.07	2.9
Lignin	College Station	7	59.86	104.4	102.2-109.3	R07020	0.32	0.14	4.8
Protein	Puerto Rico	4	58.65	39.1	34.641.3	R07020	0.13	0.21	6.9
Lignin	Puerto Rico	5	3.95	46.1	39.952.2	R07020	0.24	0.1	3.7
Xylan	Puerto Rico	5	61.53	2.8	03.2	R07020	0.17	0.08	3.3
Ash	Puerto Rico	6	58.01	85.4	82.689.4	R07018	0.28	0.16	5.1
Ash	Puerto Rico	6	58.35	92.5	89.493.5	R07018	0.21	0.13	3.7
Ash	Puerto Rico	7	4.98	49	37.254	R07018	0.19	0.09	3.3
Lignin	Puerto Rico	8	52.26	89.5	84103.7	R07020	0.28	0.1	3.5
Sucrose	Puerto Rico	8	52.26	103.5	87.3108	R07018	0.53	0.13	3.9
Glucan	Puerto Rico	9	54.79	115.8	110121.3	R07020	0.35	0.14	4.4
Xylan	Puerto Rico	10	58.87	122.2	117.1127.3	R07020	0.16	0.08	3.3
Glucan	Weslaco	1	56.66	115.3	113.2120.1	R07020	0.33	0.12	4.2
Lignin	Weslaco	1	13.61	55.7	48.861.7	R07018	0.35	0.14	4.4
Sucrose	Weslaco	1	13.61	55.7	53.360.6	R07020	0.88	0.13	4.1
Ash	Weslaco	2	74.18	171.9	169178.8	R07020	0.19	0.1	3.8

†Phenotypic contribution, noted as effect, is reflective of quantitative increase recorded in a 2m sample.

Table 15. Phenotypic and compositional trait means and standard deviations of 6 sorghum × sugarcane F_1 intergeneric hybrids and sugarcane cultivar TCP89-3505 recorded in Weslaco.

Trait	BSX0111	BSX3313	BSX4221	BSX5115	BSX7213	BSX7413	TCP89-3505	LSD‡
Brix, %	16.5	18.6	16.3	10.5	17.2	15.7	19.7	2.16
Height, cm	41	51	71	46	88	31	79	21.33
Biomass Yield, Mg/ha-1	60.3	33.8	61.5	46.7	68.9	18.8	131.3	41.76
Stalk Yield, Mg/ha-1	40.8	25.5	51.2	29	50.8	8.6	106.1	32.94
Leaf Yield, Mg/ha-1	19.5	8.3	10.3	17.6	18.1	10.2	25.2	10.43
Juice Yield, L/ha-1	13555	7733	16133	9377	12266	3466	39689	12550
Number of Stalks	29	20	20	32	25	13	20	14.07
Internode Length, cm	7	7	9	5	10	3	12	3.88
Stem Diameter, mm	26	24	26	26	21	22	30	6.51
Bagasse Moisture, %	79	64	61	80	61	78	66	12.19
Ash, %	5.4	4.6	4.6	6.3	5.3	5.7	3.9	0.96
Protein, %	3.8	3	2.6	2.3	3.4	3.1	2.4	0.57
Sucrose, %	9.9	12.5	14.1	6.3	11.5	8.6	21.1	3.98
Lignin, %	12.3	12.6	11.7	14.5	12.8	13.9	9.3	1.75
Glucan, %	30.8	30.3	33.7	31.4	32.6	28.7	32.1	1.21
Xylan, %	14.3	14.7	13.9	17.1	14.8	15.4	11.3	1.68

† Bagasse moisture percentage = 100 ~~x~~ × (1-(bagasse dry weight/ bagasse wet weight)).

‡ Least significant difference calculated for each trait at α=0.05

Table 16. Percent of calculated variances due to genotype, entry replication and other experimental factors. Repeatability indicates the likelihood of similar observed variances for genetic factors in similar individuals in Weslaco, Texas.

Trait	Repeatability	Genotype	Replication	Residual Error
Brix	0.72	84.6***	0	15.4
Height	0.66	64.3**	7.2*	28.5
Biomass Yield	0.67	66.2**	0	33.8
Stalk Yield	0.68	70.8***	0	29.2
Leaf Yield	0.54	38.4*	10.3*	51.3
Juice Yield	0.67	68.9**	0	31
Number of Stalks	0.43	25.2	0	74.8
Internode Length	0.65	63.2**	0	36.8
Stem Diameter	0.47	29.8	0	67.7
Bagasse Moisture %	0.62	53.9*	0	46.1
Ash	0.65	62.4**	0	37.6
Protein	0.67	68.9**	0	31.1
Sucrose	0.71	80.3***	0	19.7
Lignin	0.68	71.2**	0	28.8
Glucan	0.71	81.8***	0	18.2
Xylan	0.69	74.5***	0	25.5

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level.

Table 17. Calculated Pearson correlations of phenotypic traits of intergeneric hybrids.

Trait	Juice Yield (L/ha-1)	Total Biomass Yield (Mg/ha-1)	Stalk Yield (Mg/ha-1)	Leaf Yield (Mg/ha-1)	Number of Stalks	Height (cm)	Internode Length (cm)	Stem Diameter (mm)	Bagasse Moisture (%)
Brix	0.05	-0.03	0.08	-0.35	-0.41	0.16	0.28	-0.26	-0.56*
Juice Yield (L/ha-1)	-	0.93***	0.95***	0.58*	0.64**	0.44	0.42	-0.3	0.01
Total Biomass Yield (Mg/ha-1)		-	0.98***	0.77***	0.71***	0.63**	0.51*	-0.15	-0.01
Stalk Yield (Mg/ha-1)			-	0.62***	0.64**	0.65**	0.59**	-0.15	-0.16
Leaf Yield (Mg/ha-1)				-	0.73***	0.39	0.1	-0.11	0.43
Number of Stalks					-	0.06	0.22	-0.11	0.37
Height (cm)						-	0.62**	-0.18	-0.51*
Internode lgth (cm)							-	0	-0.59*
Stem Dia. (mm)								-	0.12
Bagasse Moisture (%)									-

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level.

Table 18. Correlation of intergeneric F_1 hybrid compositional traits.

Trait	Protein (%)	Sucrose (%)	Lignin (%)	Glucan (%)	Xylan (%)
Ash (%)	0.07	-0.84***	0.52*	-0.23	0.69**
Protein (%)	-	0.08	-0.38	-0.18	-0.43
Sucrose (%)		-	-0.83***	0.34	-0.89***
Lignin (%)			-	-0.35	0.92***
Glucan (%)				-	-0.19
Xylan (%)					-

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level.

Table 19. Phenotypic and compositional correlations of intergeneric F_1 hybrid agronomic traits.

Trait	Ash (%)	Protein (%)	Sucrose (%)	Lignin (%)	Glucan (%)	Xylan (%)
Brix	-0.71**	0.51	0.77***	-0.69**	-0.02	-0.83***
Juice Yield (L/ha-1)	-0.11	0.04	0.2	-0.22	0.4	-0.2
Total Biomass yield (Mg/ha-1)	-0.05	0.12	0.21	-0.19	0.44	-0.15
Stalk Yield (Mg/ha-1)	-0.22	0.05	0.36	-0.29	0.52*	-0.26
Leaf Yield (Mg/ha-1)	0.51*	0.29	-0.32	0.15	0.06	0.23
# Canes/ entry	0.26	-0.14	-0.23	0.22	0.18	0.28
Height (cm)	-0.02	0.2	0.3	-0.25	0.55*	-0.16
Internode lgth (cm)	-0.4	0.07	0.38	-0.19	0.54*	-0.21
Stem Dia. (mm)	-0.08	-0.09	-0.05	0.08	0.08	0.07
Bagasse Moisture (%)	0.54*	0	-0.67**	0.51*	-0.52*	0.48

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level.

Table 20. Agronomic traits of seventeen sweet sorghum lines containing the *iap* allele and Tx3361 evaluated in Weslaco 2012.

Entry	F1 Pedigree	Height (cm)	Days to 50% Anthesis	Brix	Uniformity	Stalk Juicy Rating
	Tx3361	121	57	17.1	1	9
12WF1943	B.05038/ Tx3361*ms3*	111	56	14.9	1	4
12WF1949	BTx3361*ms3*/ B. 05035	91	58	13.9	1	3
12WF1951	BTx3361*ms3*/ B. 05035	96	56	12.1	1	3
12WF1953	BTx3361*ms3*/ B. 05035	96	56	10.8	1	2
12WF1955	BTx3361*ms3*/ B. 05035	106	56	13.4	1	3
12WF1959	BTx3361*ms3*/ B. 05035	152	59	17.6	1	5
12WF1961	BTx3361*ms3*/ B. 05035	152	56	16.6	1	4
12WF1965	BTx3361*ms3*/ B. 05035	111	58	14.1	1	4
12WF1967	BTx3361*ms3*/ B. 05035	111	58	17.2	2	4
12WF1969	BTx3361*ms3*/ B. 05039	134	56	15.5	1	3
12WF1971	BTx3361*ms3*/ B. 05035	na‡	na§	16.9	1	4
12WF1973	BTx3361*ms3*/ B. 05035	167	56	16.3	2	3
12WF1983	BTx3361*ms3*/ B. 05035	111	56	12.4	1	2
12WF1985	BTx3361*ms3*/ B. 05035	111	56	15.4	1	4

† Indicates that juice volume was too low for brix measurement to be recorded

‡ Height data was not recorded.

§ Days from planting to flowering were not recorded.

¶ Stalk juiciness rating estimated by evaluating the cut stalk between the third and fourth internode. A 1 indicates very juicy and a 9 is dry stalk.